THE PHYSICAL SOCIETY

REPORTS ON PROGRESS IN PHYSICS

VOLUME XIV (1951)

EXECUTIVE EDITOR
A. C. STICKLAND

Price to non-Fellows £2 10*s. net* (*Postage* 1*s.*)
Price to Fellows (*one copy*) 27*s.* 6*d. net* (*Postage* 1*s.*)

Published by
THE PHYSICAL SOCIETY
1 Lowther Gardens, Prince Consort Road
London S.W.7
1951

PRINTED IN GREAT BRITAIN
BY
TAYLOR AND FRANCIS, LTD.,
RED LION COURT, FLEET ST.,
LONDON, E.C. 4

CONTENTS

Preface iv

Recent Advances in Ultra-Violet Absorption Spectroscopy. By W. C. Price . . 1

Anomalous Fine Structure of Hydrogen and Singly Ionized Helium. By W. E. Lamb, Jr. 19

New Techniques in Optical Interferometry. By H. Kuhn 64

The Diffraction Theory of Aberrations. By E. Wolf 95

The Spectrum of the Airglow and the Aurora. By A. B. Meinel . . . 121

The Microphysics of Clouds. By B. J. Mason and F. H. Ludlam . . . 147

Angular Correlations in Nuclear Reactions. By M. Deutsch 196

The Nuclear Interactions of Cosmic Rays. By G. D. Rochester and W. V. G. Rosser 227

Nuclear Effects in Atomic Spectra. By E. W. Foster 288

A Critical Survey of Ionospheric Temperatures. By N. C. Gerson . . . 316

Some Applications of Nuclear Physics in Medicine. By W. V. Mayneord . . 366

PREFACE

The *Reports on Progress in Physics* is again under the general supervision of the Papers Committee of the Physical Society for the selection of authors and subject matter for each volume of the Reports, and the executive editor of the publication is Dr. A. C. Stickland, Secretary–Editor of the Physical Society.

The members of the Papers Committee were as follows :

Chairman :—Professor R. E. Peierls.

Honorary Papers Secretary :—Dr. H. H. Hopkins.

Ordinary Members :—Mr. J. H. Awbery, Dr. M. Blackman, Mr. C. H. Collie, Professor S. Devons, Professor G. I. Finch, Dr. E. P. George, Mr. H. L. Kirke, Dr. A. C. G. Menzies, Professor D. M. Newitt, Dr. W. S. Stiles, Dr. H. Wilman and Mr. C. G. Wynne.

Again no detailed name index is included, but full bibliographical references are given alphabetically at the end of each report.

The cumulative name index for Volumes I–VII, which individually contain no name indexes, is now published; it is proposed to publish next year a cumulative index for authors and subject matter of Reports of all volumes from I to XV.

A. C. STICKLAND.
Secretary–Editor.

RECENT ADVANCES IN ULTRA-VIOLET ABSORPTION SPECTROSCOPY

By W. C. PRICE
King's College, London

CONTENTS

	PAGE
§ 1. Introduction	1
§ 2. Photoelectric spectrophotometers	2
§ 3. Light sources and photocells	3
§ 4. Double-beam recording spectrophotometers	5
§ 5. Cathode-ray presentation of spectra	7
§ 6. Ultra-violet microspectroscopy	8
§ 7. Colour translation ultra-violet microscopy	9
§ 8. Spectrophotometry in the vacuum ultra-violet	9
§ 9. Extension of the solar spectrum into the ultra-violet	11
§ 10. Sources for the vacuum ultra-violet	12
§ 11. High intensity light sources	12
§ 12. Photographic emulsions	13
§ 13. The logarithmic cam	14
§ 14. Interpretation of ultra-violet spectra	14
References	17

ABSTRACT. Recent progress in the design and construction of photoelectric spectrophotometers is reviewed. Their associated equipment, such as low-voltage hydrogen lamps, electronically stabilized power packs, ultra-violet sensitive photocells, etc., is discussed. Double-beam recording spectrophotometers and methods of cathode-ray presentation of spectra are described and an account is given of the methods of ultra-violet microspectrophotometry. Developments in the vacuum ultra-violet, solar spectroscopy, high intensity light sources, photographic emulsions, etc., are discussed and progress in the interpretation of ultra-violet absorption spectra is reviewed.

§ 1. INTRODUCTION

CONSIDERABLE progress has been made in recent years in the subject of ultra-violet spectroscopy. On the experimental side there has been the development of new detectors, new sources and improved types of instrument, while on the theoretical side the application of quantum mechanical methods to observed spectra has led in many cases to a satisfactory interpretation of the nature of the absorption bands of relatively complicated molecules. The advent of photoelectric spectrophotometry has greatly increased the precision of quantitative absorption measurements and has led to the commercial production of photoelectric spectrophotometers for purposes of chemical analysis which are now in widespread use. Some of these instruments are non-recording types, while others are double-beam recording spectrophotometers. The designs of the instruments and their auxiliary equipment, such as electronically stabilized low voltage arcs, ultra-violet sensitive photocells, etc., are of considerable physical interest. The use of reflecting microscopes has made possible the development of microspectrophotometric methods which promise to be very valuable in biophysical investigations.

The spectroscopy of the vacuum ultra-violet (below 2000 A.) has also received considerable attention, and it has been shown that the absorption of many molecules at these short wavelengths can be interpreted in terms of the excitation of outer electrons to higher and higher levels, eventually culminating in photo-ionization. Attempts have been made to use photoelectric methods in this region also, but so far with only limited success. The important part played by the absorption of short-wavelength radiation in the upper layers of the atmosphere has been recognized, and data on the ultra-violet spectra of the atmospheric gases have been used to support theories of the phenomena occurring there.

On the theoretical side a great deal of progress has been made in the interpretation of the bands arising from the fundamental chromophoric groups themselves and from the conjugation of such groups. The results have been tested by comparing calculated band intensities, vibrational patterns, states of polarization, etc., with the observed behaviour of the bands of the substances examined. Some success has been obtained, but more refined theoretical treatments and more specialized experimental data are required for the satisfactory interpretation of the spectra of the more complicated molecules. We shall try in what follows to discuss in greater detail some of the developments mentioned above.

§ 2. PHOTOELECTRIC SPECTROPHOTOMETERS

As there has been little change in photographic procedure we shall confine our attention to photoelectric instruments. These have acquired a remarkable popularity in recent years on account of their ease of application to accurate quantitative spectrophotometry. The optical density (or extinction) of a solution is obtained directly instead of being laboriously extracted via the process of photographic spectrophotometry. While many individual ultra-violet spectrophotometers had previously been constructed by different investigators (Dobson and Prefect 1930, Hogness, Zscheile and Sidwell 1937), the first commercial instrument to appear was the Beckman ultra-violet spectrophotometer (Cary and Beckman 1941). This had actually been preceded by a spectrophotometer designed by Dobson and made commercially by R. & J. Beck, but as this was a specialized instrument used for determination of the ozone content of the atmosphere we shall not consider it further. The Beckman instrument is a compact, convenient and relatively inexpensive spectrophotometer of adequate resolution and stability. Its use had led to a great increase in the application of ultra-violet absorption spectroscopy to chemical analysis and to a corresponding growth in the literature on this subject. It has been followed in England by the 'Uvispek', made by Hilger and Watts *, and a similar instrument made by Unicam. In these instruments light from the source A is focused on to the slit D of a mirror-collimated Littrow monochromator (Figure 1). The narrow band of wavelengths emerging from the monochromator then passes through the absorbing material G and falls on the photocell H. The selection of wavelengths in such a simple system is effected by rotating the prism F, usually a 30° back-aluminized quartz prism. While the use of the concave mirror E instead of lenses for collimating the beam simplifies the construction of the monochromator, the spherical and off-axis aberrations thereby introduced

* *J. Sci. Instrum.*, 1950, **27**, 109.

prevent the full resolving power of the prism from being obtained. In most cases this is not a serious drawback since, except for substances with very narrow absorption bands, the resolution is quite adequate. The Uvispek employs an on-axis system in which the prism is vertically beneath the line joining the centre of the slit to the centre of the mirror. Coma in such an arrangement is along, and not across, the slit and does not therefore have such a serious effect.

The measuring units in simple spectrophotometers employ a circuit in which a potentiometer is used to measure the potential developed by the photocell across a high resistance. The current through the potentiometer is adjusted to give a full-scale deflection with the ' cell plus solvent ' in the beam. The balance obtained with the ' cell plus solution ' moved into position gives the percentage transmission provided the dark current has previously been balanced out, the light source has remained steady and the stray light in the monochromator is of negligible proportions.

In contrast to photographic instruments such as the Hilger Spekker it will be noticed that the sample is placed after and not in front of the monochromator. This arrangement minimizes the amount of photo-decomposition that may

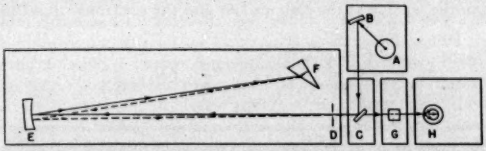


Figure 1. Schematic diagram of Beckman ultra-violet spectrophotometer.

occur in the sample though it may cause some error in working with samples which fluoresce. Ideally it would be best to use a double monochromator with the sample placed between the two halves of the monochromator, but this is difficult to achieve. Attachments are made for these instruments which enable the fluorescence or reflectance of samples to be measured at different wavelengths. These attachments fit on to the carrier slide which normally carries the absorption sample. An unknown and a standard sample can thus be rapidly compared. The main component of the attachment is a concave mirror with a hole in its centre. This is placed behind the exit slit with the reflecting surface facing away from it. Light from the monochromator passes through the hole and falls on a disc of the sample. The reflected or fluorescent light is then focused by the mirror on to the photocell. A similar arrangement with a black disc or a standard sample can be moved into the beam for comparison.

§3. LIGHT SOURCES AND PHOTOCELLS

The important developments which have made these simple spectrophotometers possible are (i) the production of electronically-stabilized ultra-violet light sources and (ii) the development of sensitive photocells in quartz or ultra-violet-transparent glass envelopes. The high-voltage hydrogen discharge tubes which had previously been employed for providing a background of continuous radiation in the ultra-violet showed fluctuations of the order of 5% in intensity, and it was therefore impossible with such sources to obtain accurate

values of the percentage transmission of a sample from successive readings by the 'sample in/sample out' method. The use of a low-voltage hydrogen arc, the current through which was electronically stabilized, gave a light source whose output was steady to a few tenths of one per cent over long periods. Another advantage of such a source is its conveniently small dimensions, the luminous discharge being largely concentrated in a short path length behind a hole in an electrode of about 2 mm. diameter. This feature enables it to be easily and efficiently focused on to the slit of the monochromator. Figure 2 shows two alternative arrangements of the electrodes in low-voltage hydrogen arcs. In the first arrangement the cathode consists of a coated filament inside a nickel box one side of which has a hole lined with a refractory metal such as molybdenum or tantalum. The box acts as a source of electrons which pass through the hole to the anode P. The discharge is localized in the electron stream between C and P and is viewed through a hole in P. The filament is sometimes mounted to one side of the box so that it is not in the direct line of sight. This minimizes fluctuations in output arising from the wandering of the arc to different spots on the filament. In the second type the box is the cathode and the anode P is a nickel plate mounted within it. An oxide-coated filament is built on to the outside of the cathode box and the electrons from this maintain

Figure 2. Types of electrode arrangement in low-voltage hydrogen arcs.

a pilot discharge to P which confines the main discharge to the region about the hole in the cathode. Tubes of both types are now available commercially in this country. Hydrogen pressures of about 3 mm. Hg are employed in the arcs and, since they run at approximately 90 volts with currents of about 0·5 amp., they can be conveniently run off electronically stabilized power packs.

Circuits suitable for the electronic stabilization of hydrogen lamps are given by Coor and Smith (1947). It is usual to reduce the filament currents to a low value once the tubes have 'struck', since bombardment with ions then suffices to maintain the filament temperature sufficiently high to give adequate electron emission. Lamps which are difficult to start can usually be made to strike if a Tesla coil is brought near them.

Lamps carrying greater currents than 0·5 amp. are not usually used in photoelectric spectrophotometry because of the difficulty in providing the necessary electronic stabilization and because higher current tubes usually require water-cooled electrodes and the continuous pumping of a stream of hydrogen through them. They are not therefore so convenient as the smaller sealed-off lamps. However, they have to be used in certain applications where great intensity is needed. High intensity low voltage arcs with water-cooled electrodes have been described by Allen (1941).* More recently Finkelstein (1950) has described a development of this tube which operates at about 14 amp. 90 volts from a 220-volt D.C. line. A lifetime of about 40 hours is quoted, this

* To obtain a long life from this lamp the filament should be turned through a right angle from the position described in the original paper.

being limited by fracture of the filament. A tube of slightly different construction, in which the filament was replaced by a hollow nickel electrode packed with oxide-coated gauze, was used by Turkevitch and Fred (1942) in their high resolution work on the rotational structure of the near ultra-violet bands of benzene. This tube was started with a high-voltage discharge, it being switched over to low-voltage operation when the cathode had become adequately heated by ion bombardment. The maximum current in the arc was about 20 amp. Lamps of the Allen type fitted with fluorite windows have been used by Dyne and Style (1947) to obtain photodissociation and fluorescence in the Schumann region. Other high intensity discharge tubes based on the discharge of a bank of condensers will be described later.

The vacuum photocells used for the ultra-violet have been made either with Corning 9741 ultra-violet transmitting glass or fused quartz envelopes and have a range extending down to 2000 A. The R.C.A. C7032 and the Cintel QVA photocells have caesium–antimony surfaces. Photomultipliers operating down to these short wavelengths are the R.C.A. 1P28 and the Mazda 27M3. Photomultipliers with 'in-blown' windows of quartz have been found to respond to the spectrum of the hydrogen arc to 1530 A. and the R.C.A. 1P28 to 1850 A. (Dunkelman and Lock 1950). By coating the window with a thin film of oil or other fluorescent material the range goes down to 900 A., and is then only limited by the emission of the hydrogen arc (Johnson and Tousey 1950). In view of the fact that photographic plates can be made sensitive down to a few Ångströms by fluorescent coatings there is no doubt that the range of photomultipliers can also be extended this way. However, it is more desirable to develop photocathodes which are only sensitive to short-wave radiation, so that the effects of scattered radiation of long wavelengths are eliminated. Suitable photocathode materials fulfilling these requirements will be discussed later.

§ 4. DOUBLE-BEAM RECORDING SPECTROPHOTOMETERS

The instruments so far described have been non-recording spectrophotometers, and an absorption curve has to be obtained by plotting a series of densities determined at a fairly large number of chosen wavelengths. With broad absorption bands this does not present great difficulty, but with narrow bands some structure can be missed unless a very large number of points are taken. Efforts have therefore been directed towards the construction of recording spectrophotometers which, operating automatically, give a pen record of the absorption curve of a sample in a relatively short time. The first of these was the conversion by Coor and Smith (1947) of an ordinary Beckman instrument to a recording spectrophotometer. A later development, along somewhat similar lines, appeared as the Cary Recording Spectrophotometer manufactured by the Applied Physics Corporation.* Light from a source which may be either a low-voltage hydrogen tube or a tungsten filament lamp, according to the wavelength desired, passes into a double monochromator, each half consisting of a Littrow mounting with a 30° quartz prism. On emerging from the exit slit the monochromator beam is divided into two halves which fall on to two 1P28 photomultipliers. In one of these beams is placed the cell plus solution and into the other a matched cell plus solvent. The outputs of the

* See *Rev. Sci. Instrum.*, 1946, **17**, 558, and *Industr. Engng. Chem.* (An. Edn.), 1947, **39**, 75 A.

photomultipliers are amplified, A.C. amplication being used and the light incident upon the monochromator being chopped at 90 c/s. They are then fed into a recording potentiometer which gives a record of the percentage transmission or, with suitable modification, a record of the densities. The drive of the monochromator is arranged so as to give a linear wavelength scale on the record and the slits are automatically opened so as to compensate for the lower intensities at short wavelengths. Compensation also has to be made for the slight differences in the variation of the sensitivities of the two photomultipliers with wavelength which causes some variation of the base line. This is a small effect, usually amounting to only about 1% over several hundred Ångströms. A record of the absorption of benzene vapour obtained with such an instrument is shown in Figure 3. The amount of fine structure present is an indication of the high resolving power of the instrument, which corresponds to a few Ångströms at a wavelength of 2500 A.

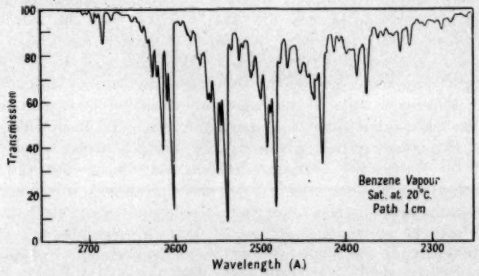


Figure 3. Record of the absorption spectrum of benzene obtained with a Cary recording spectrophotometer.

A different double-beam system has been employed by Holiday (1950) in his recording spectrophotometer. Light from the source is divided into two beams by a Hilger ' Spekker ' attachment placed in front of the monochromator. One beam passes through the cell plus solution and the other through the cell plus solvent. In the latter beam an adjustable diaphragm is located which can be used to compensate for absorption occurring in the former. A rotating sector causes first one beam and then the other to fall on the photocell, which is a Mazda 27M3 9-stage photomultiplier with fused quartz envelope. Any difference in these beams gives rise to an alternating signal which, when amplified, can be used to adjust the diaphragm until the balance is restored. A pen coupled to the diaphragm is used to record optical densities and the monochromator is geared so as to give a linear wave number scale on the record. Special electronic devices permit the spectrum to be scanned quickly where the absorption curve is fairly flat, and slow down the recording rate when discrete structure is present, thus enabling the mechanism to respond faithfully to rapid changes of absorption coefficient with wavelength. Of the two recording systems described above, the second is probably to be preferred because the first depends upon the constancy of the performance of two photomultipliers and their associated electronic equipment, while the second is independent of these and depends only upon the

calibration of a mechanical aperture. However, the fact that a large number of Cary instruments are now in operation in various laboratories is an indication of the reliability and accuracy of their performance.

There is a further system of double-beam recording which has been successfully applied to infra-red recording spectrophotometers and is equally applicable in the ultra-violet. This is embodied in the ratio-recording double-beam spectrophotometer of Savitsky and Halford (1950). The system makes use of the phase discrimination characteristics of the breaker or commutator rectifying system. The chopping of the two beams which fall on a single detector takes place not alternately but 90° out of phase. The output of the amplifier passes through a phase inverter and impedance matching system to a commutator geared to the beam chopper. A pair of breakers at 90° to one another can then be set so that each produces a D.C. component due to the signals from one of the beams and an A.C. component due to the other, which can be removed by filters. The outputs from the two breakers are then fed into the recording potentiometer, which thus gives a record of the percentage transmission.

Another interesting device which is being produced by Hilger is a photoelectric attachment which can be used on their ' Medium ' spectrograph alternatively with the photographic plate holder, thus making it possible to combine in one instrument the quantitative accuracy of photoelectric with the high resolution of photographic methods. Advantage is taken of the flat field of this instrument to mount a photomultiplier–slit combination on a carriage driven by a screw whose axis is parallel to the focal plane. The slit moves in the focal plane of the spectrograph and so scans the spectrum. The sample is placed in front of the spectrograph as in the ' Spekker ' photometer. The optics of the spectrograph (with its on-axis lenses) are extremely good and very high resolution can be obtained, while mere removal of the photoelectric attachment and the insertion of the photographic plate holder permits its use for emission analysis or exploratory absorption work by photographic methods. At present single-beam operation is employed, but electronic elaboration could be used to convert it to a double-beam instrument.

Further details on ultra-violet spectrophotometers will be found in the *Proceedings* of the London Conference on Optical Instruments (1950) in the articles by Ballard and Thompson.

§ 5. CATHODE-RAY PRESENTATION OF SPECTRA

In order to study chemical changes which occur very rapidly it is sometimes convenient to obtain spectral records on a cathode-ray tube as a trace of transmission (or emission) against wavelength. Methods of doing this in the infra-red have been described by Daly and Sutherland (1947). The rotation of a prism or mirror is used to pass the spectrum across the detector, the motion being converted by means of a potentiometer into deflection in the X direction, while the output of the detector after suitable amplification is put on the Y plates. In the ultra-violet the optical arrangement is similar except that a photomultiplier is used instead of a bolometer. Because of the shorter response time of the former much higher chopping speeds and simpler amplifying systems may be used (Deal, Bradshaw and Matsen 1948). More recent developments using an image orthicon permit the wavelength scanning to be done electronically instead of mechanically (Benn, Foote and Chase 1949), thus correspondingly shorter

scanning times can be achieved. The image orthicon R.C.A. Type 2P23 was used in their work. This has a high luminous sensitivity in the range 3500–10,000 A. It is claimed that for short exposure times the orthicon is much more sensitive to light of low intensity than any photographic emulsion. It is sensitive not only in the visible region but also in the infra-red, and its range could be extended to 2000 A. by making the envelope of suitably transparent glass. Its response exhibits over short times a cumulative character which for short exposures to weak radiation may be more linear than is the case for a photographic plate. The application was originally made to study a gaseous combustion process in which a few prominent spectral lines and bands were available. The time variation of the intensities of these lines and bands, though unobtainable on a moving plate or film because of low intensities, was observed with ease and accuracy with the orthicon. While there appears to be a future for this tube in spectroscopy, it should however be mentioned that the trace of the mercury spectrum shown in the article by Benn, Foote and Chase was somewhat lacking in resolution.

§6. ULTRA-VIOLET MICROSPECTROSCOPY

Investigations involving ultra-violet microscopy and micro-absorption spectrophotometry have been reviewed up to 1948 by Scott and Sinsheimer (1950) and more recently by Loofbourow (1950) and Randall (1950). Micro-spectroscopy involves the combination of microscope and spectroscopic optics for the purpose of investigating the spectra of very small samples or small portions of larger ones. The types of specimen that may be observed include single crystals, minute amounts of liquids or solutions, thin solid films, biological cells and tissues. With modern microscope attachments the weight of material required for obtaining micro-absorption, emission or fluorescent spectra may be reduced by factors of up to 10^{-3} or 10^{-4} as compared with the corresponding macro-methods. This is clearly of great importance in micro-analytical procedures. It may be illustrated, for example, by the fact that in ultra-violet microscopic work quantitative estimations are often made on as little as one thousandth of a microgram of material. This is a result of a fortunate combination of the strong selective absorption of many materials in the ultra-violet and the high resolving power of the microscope attachment when used at short wavelengths. In the infra-red, where absorption coefficients are not so high and wavelengths about twenty times as great, the smallest quantity of material that can be dealt with is about a microgram.

Ultra-violet and infra-red polarization studies with small crystals and solid films promise to be useful in studying crystal structure, molecular orientation, and in investigating the solid state. Ultra-violet microscopy has made it possible to examine the distribution of important biochemical substances within biological cells or tissues in normal and pathological conditions with important results in biophysics.

One of the first to apply microspectroscopic methods in the ultra-violet was Caspersson (1936), who added a sector photometer and photocell to a Zeiss quartz microscope illuminated by a quartz monochromator. He showed that it was possible with such an arrangement to determine quantitative ultra-violet absorption spectra of small portions of biological cells and tissues. Rapid advances have recently been made as a result of the design of microscope objectives

of the reflection type which are achromatic over the whole spectral range (Brumberg 1943, Burch 1947, Wilkins and Seeds 1949, Norris, Seeds and Wilkins 1950), and as a result of continued improvement in radiation detecting devices, monochromators, light sources, and other components. A diagrammatic sketch indicating the main features of the design of a reflecting microscope objective is given in Figure 4. The method of microspectrophotometry mainly employed involves the use of a monochromator to illuminate the microscope. Focusing and selection of the part of the image to be studied is done with visible light by means of a mirror which deflects the beam, normally incident upon the photoelectric detector, through an eyepiece. This mirror is then removed from the beam and the monochromator set to the wavelength at which quantitative

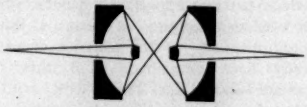


Figure 4. Schematic diagram of a reflecting microscope

transmission measurements are required. Because of the reflecting nature of the microscope optics the image remains in focus in spite of the change in wavelength. When, in order to cover a larger field, photographic methods are used, the photographic plates are calibrated *in situ* by means of a density wedge and are subsequently microphotometered (Davies and Walker 1950). This arrangement has the advantage that it cuts down to a minimum the quantity of ultra-violet radiation employed to make the observations, thus reducing to as small a value as possible the photo-decomposition and other effects of irradiation which are likely to occur. Even so, the ultra-violet dosage required is such as to limit appreciably the amount of work which can be done on living material with this technique.

§7. COLOUR TRANSLATION ULTRA-VIOLET MICROSCOPY

In the colour translation method of ultra-violet microscopy, which was originally suggested by Brumberg (1939, 1941) and recently perfected by Land and collaborators (1949), three wavelengths in the ultra-violet are selected and converted by intermediate photographic and projection steps into three visible colours. The microscopic field appears as a multicoloured projection image in which variations in hue correspond to differences in absorption at the selected ultra-violet wavelengths. Thus the chemical constitution at any position in the image can be seen as a function of the colour of the image at that point. Small structures which, because of their transparency, are not visible with ordinary microscopes, are made clearly observable without the use of dyes or stains. The method is being developed commercially by the Polaroid Corporation. It incorporates high-speed methods of film processing, so that within 10 seconds a stable negative is obtained which is suitable for projection.

§8. SPECTROPHOTOMETRY IN THE VACUUM ULTRA-VIOLET

The natural extension of the techniques of ultra-violet absorption spectrophotometry into the shorter wavelength region below 2000 A. is at present engaging the attention of many workers. While much has already been done by means of photographic methods in this range, the use of photoelectric

techniques still presents many problems. Firstly, there is the difficulty in using the usual type of photocell in this region because of the opacity of the material of the envelope. Thin fused quartz does not transmit much below 1750 A., and while synthetic lithium and calcium fluoride crystals are transparent to about 1200 A., it is not possible to outgas a photocell made with such a window because cements capable of standing the temperatures required are not available. In any case, prolonged irradiation tends to make these windows opaque, though this is probably not a serious difficulty if the intensity of the light is kept low. The use of photoconductive cells of cadmium sulphide or other materials which need not be put in a vacuum envelope or covered with a window seems attractive as the light could then be allowed to fall directly on the photo surface. However, as the sensitivity of these surfaces appears to depend upon adsorbed oxygen, they cannot in fact be used in a vacuum. The use of fluorescent coatings over the windows of photomultipliers for the purpose of extending the range of sensitivity has already been described. For this purpose the E.M.I. photomultiplier (Sommer and Turk 1950) with semi-transparent cathode evaporated directly on to the window of the tube is probably the most satisfactory design.

The construction of vacuum ultra-violet monochromators has been described by Baker (1938), Little (1946) and Parkinson and Williams (1949). They have been used for the study of photoelectric effects at short wavelengths, for the transmission and reflectance of various substances and for the determination of the position and character of energy levels which are of importance in the study of processes in luminescent materials, etc. Little use has yet been made of this region of the spectrum for quantitative absorption spectrophotometry and, except in special cases, it is doubtful whether it can be utilized for such a purpose in the same way as the longer wavelength region.

The monochromator of Parkinson and Williams employs a 1-metre concave grating ruled with 15,000 lines/inch mounted in a normal incidence mounting, the scanning of the spectrum being accomplished by the rotation of the grating. As a source they use a high-voltage hydrogen tube, and as detector a 931–A or 1P21 photomultiplier in front of which is placed a layer of manganese-activated willemite. For investigating photoconducting states with the apparatus an alternating voltage is applied across the sample, and this is modulated by irradiation of the sample with pulsed light from the hydrogen source. The alternating voltage eliminates polarization effects, and the dark current through the sample is eliminated by measuring only the photocurrent with the frequency of the pulsations.

The monochromator of Little, which is a development of one made by Powell (1934), is very similar to that of Parkinson and Williams. It was used largely for measuring reflectances and transmittances of polished specimens at various angles of incidence. Measurements of reflectance at two or more angles of incidence permit the calculation of the index of absorption and the index of refraction of materials opaque to these short wavelengths. An interesting feature of the equipment was the use of a phototube with a platinum cathode. This was fitted with a fluorite window and filled with argon. The sensitivity even with the gas filling was rather low. It was found to be greatly increased by maintaining a glow discharge in hydrogen through the tube for about 15 hours. This cleaned the platinum surface thoroughly, and the increased sensitivity enabled the tube

to be operated at voltages well below its breakdown voltage, where its characteristic was very nearly linear. The advantage of a platinum photocathode is that it does not respond to scattered visible or near ultra-violet light because the photoelectric threshold of clean platinum is about 2200 A. It was thus possible to install glass windows in the apparatus and to work in daylight. Much interest has recently been shown in photoelectric monochromators, as evidenced by several reports read at the March 1950 meeting of the Optical Society of America.* From these it would appear that the greater sensitivity of photomultipliers with fluorescent coatings is tending to outweigh the other advantages of the platinum photocell. In the instrument described by Tousey, Johnson, Richardson and Toran (1950) provision is made for the adjustment of slits without breaking the vacuum, these being carried on turrets similar to large stopcocks. It is also possible to replace the exit slit by a cassette for photographic work.

§9. EXTENSION OF THE SOLAR SPECTRUM INTO THE ULTRA-VIOLET

The great importance of the effects of ultra-violet solar radiation on the upper atmosphere has led to attempts to photograph the sun's emission at great altitudes where the effects of atmospheric absorption may be expected to be somewhat reduced. At sea level observations of the sun's spectrum are limited on the short wavelength side to about 3100 A. because of the absorption of ozone, the main bulk of which lies at an altitude of from 20 to 40 km. While it is clear that balloon or aeroplane ascents are inadequate to carry spectrographic equipment above the ozone layer, the recent use of V2 rockets capable of ascending to heights of up to 160 km. has made this possible. The spectrograph used embodied a 15,000 lines/inch grating of 50 cm. radius of curvature built into the war-head of the rocket. Accounts of the results of different rocket ascents are given by Hulburt (1947), Baum *et al.* (1946), Durand, Oberly and Tousey (1947) and Hopfield and Clearman (1948). The first trials showed that most of the atmospheric ozone layer was penetrated at between 20 and 34 km., but even at 55 km. there was still some absorption at the centre of the Hartley band (*c.* 2550 A.). Later flights attained peak heights of up to 155 km. Records from these showed extension of the solar spectrum to 2300 A., traces being evident at 2230 A. Numerous Fraunhofer absorption lines are present in this region. Of these the lines of FeI and FeII are dominant and clearly contribute the major number of lines. In addition the following lines were definitely identified: Sir, CI, MgII (also in emission), VI, VII, CrII, MnII, CoI, AlI. Lines of NaI, NiI, NiII, CrI, CoII, BeI, AlI, PI and CaI were identified with less certainty. Determinations of the nature of the black-body curve of the sun's radiation in this wavelength range are difficult. There are a number of sudden drops in intensity—e.g. near 2630, 2550 and 2410 A.—due to multiplets of iron. There are step-wise drops of intensity, such as one at 2950 A., that might originate in the reversing layer of the sun, also several bright windows, e.g. bright lines near 2638 and 2643 A. It appears, however, that, even taking the bright windows, they are less bright than should be expected for a solar temperature of 5,500° K. The difficulty at present is to obtain sufficiently good 'homing' on the sun to obtain strong spectra. When this is achieved, and also greater heights attained, then

* *J. Opt. Soc. Amer.*, 1950, **40**, 259.

it might be possible to extend observations to 1000 A. or less, i.e. to the radiation which is responsible for conditions in the ionosphere, and this would be of very great interest indeed.

§ 10. SOURCES FOR THE VACUUM ULTRA-VIOLET

No new sources giving a continuous background in the vacuum ultra-violet have been developed. The hydrogen, the helium (Hopfield 1930, Henning 1932) and the Lyman continuum are still the only three we have. However, there have been several developments in connection with Lyman sources aimed at replacing worn-out capillaries easily and without disturbing the spectrographic alignment (Worley 1942, Maeder 1948, Collins and Price 1934). As this source is the only one which covers the whole range from 2000 A. down to a few hundred Ångströms it would be most valuable if it could be improved or replaced.

§ 11. HIGH INTENSITY LIGHT SOURCES

A light source of high intensity in the ultra-violet has been developed by Porter and Norrish (Norrish and Porter 1949, Porter 1950). This involves the triggered discharge of large banks of condensers of up to 1,000 μF. capacity charged to 4,000 volts through a quartz tube about 1 metre long containing a rare gas at a few centimetres pressure. The flash obtained, though only lasting for a few milliseconds, has an intensity of up to ten thousand times that obtainable from ordinary light sources. A single flash is capable of producing a photo-decomposition of over 80% in a gas at several centimetres pressure in an absorption tube placed alongside the source. Radicals are produced in sufficiently high concentration for their absorption spectra to be photographed under high dispersion. The high rate of photolysis naturally results in quite different products from those obtained with low intensity sources, and can perhaps be compared with the photo-dissociation that may be expected to occur in the vicinity of atomic explosions.

Small concentrated sources of high intrinsic brilliance are desirable for micro-spectroscopy in order to cut down exposure times. Sources based on the hydrogen continuum have already been described. It has been more customary up to the present, however, to use high wattage mercury-in-quartz arcs or, alternatively, sparks between rotating cadmium electrodes. Commoner (1950), for example, uses a water-cooled high pressure 1,000-watt mercury capillary lamp (General Electric Type AH6), and Davies (1950) uses a Thermal Syndicate 125-watt medium pressure mercury arc stabilized in position by two constrictions in the quartz envelope. Land, in his colour translation microscope, uses a mercury arc similar to that described by Noel (1940). A description of the characteristics of these high-pressure mercury arcs is given by Harrison, Lord and Loofbourow (1948). Because of its high radiant emission from a small emitting area the G.E.C. H–6 arc is an excellent source for applications in which high radiance is required in the range from about 2700 A. through the visible and near infra-red regions and in which a strong continuous spectrum is desirable. A similar lamp, made by Phillips, has been used by Caspersson and his co-workers. These lamps have the disadvantages that they are expensive and have a high breakage rate. For shorter wavelengths, or for the isolation of monochromatic radiation, other sources may be preferable. Compact source lamps of the type described by Bourne (1945), where the mercury pressure is relied upon to confine the arc to a small gap between tungsten electrodes, are not suitable for short-wave work. This is because, when

run at the full rated wattage, their emission does not usually extend to wavelengths shorter than about 3400 A. on account of broad reversal by the long path of mercury vapour to be traversed. Some use at short wavelengths can, however, be made of these and other commercial lamps in quartz envelopes if they are run at about half the normal wattage. Cadmium lamps in quartz envelopes have not yet proved very successful owing to the preliminary heating required to vaporize the cadmium, and also to the fact that the 2573 and 2748 A. lines, which are the useful ones, being spark lines, are weak in the arc. The old type spark between rotating electrodes has many disadvantages in terms of noise and the fumes given off (Barnard and Welch 1936). Other new sources, such as the zirconium concentrated-arc lamp (Buckingham and Diebert 1946) and the gas arc (Aldington 1949), though very intense in the visible, do not emit a useful amount of ultra-violet radiation. Extended source high wattage low vapour pressure mercury lamps with water-cooled electrodes suitable for ultra-violet Raman spectroscopy, fluorescence or photochemical work have been described by Rank and Kagarise (1950).

§ 12. PHOTOGRAPHIC EMULSIONS

The development of commercial photographic emulsions of the Schumann type in recent years has been a great boon to spectroscopists working in the region below 2000 A. In England, Ilford Q emulsions, which can be coated on very thin glass for use with short radius gratings, have proved very satisfactory in the region from 2000–100 A., though they do not have quite the resolving power of the best Schumann plates. The Q emulsion was originally produced in research on the photographic recording of positive rays. The characteristics of a similar emulsion recently developed by Kodak, known as Kodak Short Wave Radiation emulsion (abbreviated S.W.R.) are described by Schoen and Hodge (1950). Both these emulsions are greatly superior to those employing fluorescent coatings because of their greater resolving power and their insensitivity to long wavelength scatter. They contain a minimum of gelatine and the silver halide grains poke out through the surface of the supporting film, and are thus exposed to the direct action of the light. For this reason the emulsions are much more easily damaged by abrasion than emulsions in which the halide grains are protected by being embedded within the gelatine.

Work has also been carried out on the most suitable emulsions for special purposes in the 2000–4000 A. range. When plenty of light is available, or the sample being examined is unaffected by long exposure, the choice of emulsion to give the highest resolving power is fairly straightforward. However, when a compromise between speed and graininess has to be effected, as in the photomicrography of living material, the choice is more difficult. This aspect has been dealt with by Fraser (1950), who suggests a criterion of merit for the selection of suitable emulsions. This index of quality is taken as $100\,S/G^2$, where S and G are respectively the speed and grain size relative to the Kodak B.20 Process Regular Plate. Measured values of S and G for various ciné-film material in the region 2650 A. are reported in Fraser's paper. They have been used, for example, by Davies (1950) in choosing the most suitable emulsion for the serial photomicrography of cells at different ultra-violet wavelengths. It is clear that the index of quality is also of importance in photographing the spectra of weak sources of the ultra-violet.

§ 13. THE LOGARITHMIC CAM

In order to bring out the structure present in some diffuse bands a device due to Holiday (1937) has proved very successful. This consists of a logarithmic cam which is coupled to the plate holder of a spectrograph. It produces a logarithmic displacement of the photographic plate during the exposure, and it has the effect of greatly enhancing the appearance of small variations in intensity of absorption which otherwise might be completely overlooked. Using this device, Brown and Randall (1949) were able to reveal considerable structure in cooled films of biologically important materials.

§ 14. INTERPRETATION OF ULTRA-VIOLET SPECTRA

So far in this report we have only discussed developments on the instrumental side of ultra-violet absorption spectroscopy. To deal with the interpretative side in any detail, especially with its implications concerning the molecular structure of polyatomic molecules, would require a complete volume. Fortunately there exist many excellent review articles and books on the various aspects of ultra-violet spectra, which also serve as keys to the literature on the subject. In what follows we shall mention briefly the contents of some of these reviews so that a broad picture of the subject will be obtained, and any particular aspect of it can then be followed up in a specialized review article. With regard to the general outline of the subject there is firstly the empirical use of absorption spectra for the purpose of quantitative spectrochemical analysis—inorganic, organic and biological. Then there is the physico-chemical interest in the spectra from the point of view of molecular structure, energy levels, chemical activity, theory of colour, etc. This varies in treatment from a semi-empirical approach to more rigorous attacks by quantum-mechanical methods. The value of the data is important in that they enable quantum theoretical predictions to be tested and special mathematical methods to be developed which are of fundamental importance to a proper understanding of the physical and chemical properties of matter. The theoretical treatment of absorption spectra involves methods of representing and expressing the electronic configuration of molecules in their normal and excited states, of deriving the selection rules which govern the transitions which occur, and of calculating the frequencies and intensities of the associated bands. Ultra-violet spectra provide information which is complementary to that given by infra-red and Raman spectra, x-rays, electron diffraction, etc., but it differs from them in dealing more directly with the electrons rather than with atomic vibrations or interatomic distances which are the consequences of forces derived from the nature of the electron configuration.

First we shall mention briefly the quantitative aspect of spectrophotometry. Several good books are now available which deal with this and with the empirical correlation of near ultra-violet absorption spectra with molecular structure (Morton 1942, Brode 1939, Lothian 1949). What is needed in the ultra-violet as well as in the infra-red is as comprehensive a collection of curves as possible or, less preferably, an index to the literature to enable these to be found quickly. All that is available at the present time are the collections of curves published by the American Petroleum Institute, which are mainly on aromatic hydrocarbons, and the out-of-date collections in the *International Critical Tables* and the Landolt–Bornstein Tables. The number of articles giving absorption data is very numerous, but they are so widely scattered through the literature that it is sometimes difficult

to trace them. Valuable literature references are given in the books already referred to and in many of the review articles mentioned later. Annual reviews are usually published in *Analytical Chemistry* (Rosenbaum 1950) or in the *Annual Reports of the Chemical Society* (see also Marrison 1947, 1948). A comprehensive bibliography of the spectra of diatomic molecules up to 1949 is given by Herzberg (1950).

The correlation of a vast amount of experimental data on the absorption of organic molecules with their chemical structure is reviewed by Morton (1941), Braude (1945), Maccoll (1947) and other articles to which they refer. Introductions to the fundamental theory of the electronic structure and spectra of polyatomic molecules are given in reports by Bowen (1942, 1943), Coulson (1947), Walsh (1947, 1948) and Price (1939). Rigorous treatment by the methods of group theory is excellently presented by Sponer and Teller (1941) and Sponer (1942). The spectroscopy of organic molecules in the vacuum ultra-violet has been reviewed by Platt and Klevens (1941) and by Walsh (1947). Numerous articles have appeared in the last two years on the theory of the electronic structures of molecules, particularly aromatic molecules, by Mulliken, Coulson, Longuet-Higgins, Craig, Platt, Bayliss, Kasha, etc. These are concerned with the derivation of the character and energies of the excited states of molecules and the interpretation of the associated spectra. The September (1950) *Discussion of the Faraday Society* contains fifteen important papers on ultra-violet absorption spectra, one of which reviews the present position of the subject (Coulson 1950). A previous Report by Mulliken and Rieke (1941) was concerned with the theoretical interpretation of oscillator strengths and intensities and their relation to ultra-violet spectra, dispersion and polarization.

Symposia dedicated to the memory of two eminent workers in the field—Victor Henri and G. N. Lewis—were held in 1947. The proceedings of the first is published in a volume entitled *Contribution a l'étude de la structure moléculaire* (Liége: Desoer, 1948). The second, on *Colour and the Electronic Structure of Complex Molecules*, appears in *Chemical Reviews*, 1947, **41**. Both contain many valuable articles relevant to this report. The work of G. N. Lewis and his associates led to the identification of the phosphorescent state with the lowest triplet state of a complex molecule. The total luminescence of a complex molecule in a rigid glassy medium (e.g. supercooled solutions in alcohol/ether/isopentane mixture) is usually found to occur in two spectral regions. The first occurs at shorter wavelengths and is readily identified by its spectral position and short emission lifetime as the reverse of the lowest normal absorption process. The longer wavelength luminescence has a relatively long mean lifetime, and spectroscopic and magnetic studies indicate that the associated metastable state is the lowest triplet state of the molecule. For instance, the longer wavelength luminescence in benzene coincides with an extremely weak absorption at about 3400 A., for which the upper state is predicted as the lowest triplet state of the molecule. The short wavelength fluorescence is the same electronic transition as the A_{1g}–B_{2u} absorption band at 2600 A., and has a correspondingly short emission lifetime.

To conclude this report we shall discuss briefly some of the work done on photo-ionization in the vacuum ultra-violet. We choose this from the numerous other aspects of molecular spectroscopy we might discuss partly because of the precision it has attained in the measurement of ionization potentials and partly

because, as a region of the spectrum, the vacuum ultra-violet is less well known than the more accessible regions at longer wavelengths. It is important also because the strongest absorption bands of simple substances like water, the halogen acids, acetylene, ethylene saturated hydrocarbons, aldehydes, etc., are to be found in the region below 2000 A. Absorption at longer wavelengths than this usually only occurs in relatively complex molecules, but even these molecules also have strong bands ($\log\epsilon \sim 4$ to 5) in the vacuum ultra-violet. The reason for this strong absorption is that the ionization potentials of the outer electrons of most molecules and atoms lie between 6 and 15 electron volts, corresponding to photons of wavelengths approximately 2000 A. and 500 A. respectively. Irradiation with light of these wavelengths will give rise to a set of absorption bands corresponding to changes of principal quantum number. These will form a series of electronic bands similar to an atomic Rydberg series and corresponding to the resonance potentials of the molecule. On going to short wavelengths the bands get closer together and eventually fuse into a continuum which represents the photo-ionization of the molecule. The extrapolation of the higher members of the series by means of a Rydberg formula yields a very accurate value of the ionization potential of the molecule. A good example of such a set of bands is the absorption spectrum of methyl iodide (Figure 5). Here two

Figure 5. The absorption spectrum of methyl iodide in the region 1400–1220 A.

such series are obtained, and they can be interpreted as leading to the doublet state of the molecular ion ($^2\Pi_{3/2}$ and $^2\Pi_{1/2}$ in simplified notation). The excitation and photo-ionization of inner groups of electrons have been observed for simple molecules like oxygen, nitrogen, nitric oxide, carbon dioxide (Price and Collins 1935, Worley and Jenkins 1938, Tanaka 1942, Henning 1932), but because of the shorter wavelengths at which the spectra occur, and for other reasons, they are usually more difficult to obtain. The outer electrons of a molecule are generally non-bonding electrons, and in the case of methyl iodide these are the $5p\pi$ iodine electrons. The removal of these does not produce much molecular distortion, and little, if any, vibrational structure accompanies the main electronic bands. For molecules such as ethylene or acetylene the electrons of the lowest ionization potentials are the π electrons of the multiple bond, and naturally it is to be expected that the removal of these electrons will result in vibration in this bond. The successive electronic states found for these molecules are in fact represented by short vibrational patterns, the main separation of which corresponds to the valence vibration of the multiple bond in the excited state. The study of the spectra is clearly capable of giving much information about the electrons which constitute these important chemical bonds, which are the basis of much of organic chemistry. For example, the rotational structure of the bands of acetylene was the first direct spectroscopic evidence that the electrons concerned were of a π type.

The accumulation of a large number of accurate spectroscopic ionization potentials of various molecules is providing a fund of data which is greatly increasing our knowledge of the electronic structure of molecules. The effect of alkyl substitution has been shown to reduce appreciably the ionization potentials of the electrons of bonds to which they are attached. This is interpreted as a stabilization of the ionic state by the polarization of the attached group (Price 1947). This gives us a measure of how much an excited state may be stabilized, or an energy of activation reduced, by bringing polarizable material into the vicinity of a chemical bond, and this is of fundamental importance to chemical reactivity. The study of the reduction of the ionization potential with increasing conjugation is also of great interest. In going from ethylene to butadiene, or benzene to naphthalene, the ionization potentials fall from 10·50 to 9·07 or 9·24 and to 8·1 v. respectively as a result of a process which is somewhat analogous to the drop from the ionization potential of, say, a free sodium atom to the photo-electric threshold of a film of the metal. Finally, it should be mentioned that the photo-ionization of the atmospheric gases by short-wave ultra-violet light is responsible for the various layers of the ionosphere, and information which can be obtained about the processes of such ionization is of great value to the theories of the effects occurring in these regions.

REFERENCES

ALDINGTON, J. N., 1949, *Trans. Illum. Engng. Soc., Lond.*, **14**, 9.

ALLEN, A. J., 1941, *J. Opt. Soc. Amer.*, **31**, 268.

BAKER, R. F., 1938, *J. Opt. Soc. Amer.*, **28**, 55.

BALLARD, S. S., 1950, *Proceedings of London Conference on Optical Instruments* (in the press).

BARNARD, J. E., and WELCH, F. V., 1936, *J. Roy. Microscop. Soc.*, **56**, 365.

BAUM, W. A., JOHNSON, F. S., OBERLEY, J. J., ROCKWOOD, C. C., STRAIN, C. V., and TOUSEY, R., 1946, *Phys. Rev.*, **70**, 781.

BENN, R. E., FOOTE, W. S., and CHASE, C. T., 1949, *J. Opt. Soc. Amer.*, **39**, 529.

BOURNE, M. K., 1945, *J. Sci. Instrum.*, **22**, 107.

BOWEN, E. J., 1942, *The Chemical Aspects of Light* (Oxford: Clarendon Press), p. 1; 1943, *Annual Rep. Chem. Soc.*, **40**, 12.

BRAUDE, E. A., 1945, *Annual Rep. Chem. Soc.*, **42**, 105.

BRODE, W. R., 1939, *Chemical Spectroscopy* (New York: John Wiley and Sons).

BROWN, G. L., and RANDALL, J. T., 1949, *Nature, Lond.*, **163**, 209.

BRUMBERG, E. M., 1939, *C.R. Acad. Sci. U.R.S.S.*, **25**, 473; 1941, *Ibid.*, **32**, 486; 1943, *Nature, Lond.*, **152**, 357.

BUCKINGHAM, W. D. and DEIBERT, C. R., 1946, *J. Soc. Mot. Pict. Engrs.*, **47**, 376.

BURCH, C. R., 1947, *Proc. Phys. Soc.*, **59**, 41.

CARY, H., and BECKMAN, A. O., 1941, *J. Opt. Soc. Amer.*, **31**, 682.

CASPERSSON, T., 1936, *Skand. Arch. f. Physiol. Supplement*, **73**.

COLLINS, G. B., and PRICE, W. C., 1934, *Rev. Sci. Instrum.*, **5**, 423.

COMMONER, B., 1950, *Discussions of the Faraday Soc.*, in the press.

COOR, T., and SMITH, D. C., 1947, *Rev. Sci. Inst.*, **18**, 173.

COULSON, C. A., 1947, *Quart. Rev. Chem. Soc.*, **1**, 144; 1950, *Discussions of the Faraday Soc.*, in the press.

DALY, E. F., and SUTHERLAND, G. B. B. M., 1947, *Proc. Phys. Soc.*, **59**, 77.

DAVIES, H. G., 1950, *Discussions of the Faraday Soc.*, in the press.

DAVIES, H. G., and WALKER, P. M., 1950, *Discussions of the Faraday Soc.*, in the press

DEAL, W. E., BRADSHAW, W., and MATSEN, F. A., 1948, *J. Chem. Phys.*, **16**, 638.

DOBSON, G. M. B., and PREFECT, D. S., 1930, *Photoelectric Cells and their Applications* (London: Physical Society), p. 1.

DUNKELMAN, L., and LOCK, C., 1950, *J. Opt. Soc. Amer.*, **40**, 264.

DURAND, E., OBERLEY, J. J., and TOUSEY, R., 1947, **71**, 827.

DYNE, P. J., and STYLE, D. W. G., 1947, *Discussions of the Faraday Soc.* A, **4**, 159.

FINKELSTEIN, N. A., 1950, *Rev. Sci. Instrum.*, **21,** 509.
FRASER, R. D. B., 1950, *J. Sci. Instrum.*, **27,** 106.
GREY, D. S., 1950, *J. Opt. Soc. Amer.*, **40,** 283.
HARRISON, G. R., LORD, R. C., and LOOFBOUROW, J. R., 1948, *Practical Spectroscopy*, (New York : Prentice-Hall, Inc.; London : Blackie), p. 186.
HENNING, H., 1932, *Ann. Phys., Lpz.*, **13,** 599.
HERZBERG, G., 1950, *Molecular Spectra and Molecular Structure : I—Spectra of Diatomic Molecules*, 2nd Edition (New York : Van Nostrand), p. 501.
HOGNESS, T. R., ZSCHEILE, A. E., and SIDWELL, F. P., 1937, *J. Phys. Chem.*, **41,** 371.
HOLIDAY, E. R., 1937, *J. Sci. Instrum.*, **11,** 166; 1950, *M.R.C. Spectrographic Unit, London Hospital, E.*1, details to be published.
HOPFIELD, J. J., 1930, *Astrophys. J.*, **72,** 133.
HOPFIELD, J. J., and CLEARMAN, H. E., 1948, *Phys. Rev.*, **73,** 877.
HULBURT, E. O., 1947, *J. Opt. Soc. Amer.*, **37,** 405.
JOHNSON, F. S., and TOUSEY, R., 1950, *J. Opt. Soc. Amer.*, **40,** 264.
LAND, E. H., BLOUT, E. R., GREY, D. S., FLOWER, M. S., HUSEK, H., JONES, R. C., MATZ, C. H., and MERRILL, D. P., 1949, *Science*, **109,** 371.
LITTLE, E. P., 1946, *J. Opt. Soc. Amer.*, **36,** 168.
LOOFBOUROW, J. R., 1950, *J. Opt. Soc. Amer.*, **40,** 317.
LOTHIAN, G. F., 1949, *Absorption Spectrophotometry* (London : Hilger and Watts, Ltd.).
MACCOLL, A., 1947, *Quart. Rev. Chem. Soc.*, **1,** 16.
MAEDER, R., 1948, *Helv. Phys. Acta*, **21,** 411.
MARRISON, L. W., 1947, *Industr. Chem. (Lond.)*, p. 817; 1948, *Ibid.*, p. 146.
MORTON, R. A., 1941, *Annual Rep. Chem. Soc.*, **38,** 7; 1942, *Application of Absorption Spectra to Vitamins, Hormones and Co-Enzymes* (London : Hilger and Watts, Ltd.).
MULLIKEN, R. E., and RIEKE, C., 1941, *Rep. Prog. Phys.*, **8,** 251 (London : Physical Society).
NOEL, E. B., 1940, *J. Appl. Phys.*, **11,** 325.
NORRIS, K. P., SEEDS, W. E., and WILKINS, M. N. F., 1950, *J. Opt. Soc. Amer.*, in the press.
NORRISH, R. G. W., and PORTER, G., 1949, *Nature, Lond.*, **164,** 658.
PARKINSON, W. W., and WILLIAMS, F. E., 1949, *J. Opt. Soc. Amer.*, **39,** 705.
PLATT, J. R., and KLEVENS, H. B., 1941, *Rev. Mod. Phys.*, **16,** 182.
PORTER, G., 1950, *Proc. Roy. Soc.* A, **200,** 284.
POWELL, W. M., 1934, *Phys. Rev.*, **45,** 154.
PRICE, W. C., 1936, *J. Chem. Phys.*, **4,** 539; 1939, *Ann. Rep. Chem. Soc.*, **36,** 47; 1947, *Chem. Rev.*, **41,** 257.
PRICE, W. C., and COLLINS, G. B., 1935, *Phys. Rev.*, **48,** 714.
RANDALL, J. T., 1950, *Discussions of the Faraday Soc.*, in the press.
RANK, D. H., and KAGARISE, R. E., 1950, *J. Opt. Soc. Amer.*, **40,** 89.
ROSENBAUM, E. J., 1950, *Analyt. Chem.*, **22,** 14.
SAVITSKY, A., and HALFORD, R. S., 1950, *Rev. Sci. Instrum.*, **21,** 203.
SCHOEN, A. L., and HODGE, E. S., 1950, *J. Opt. Soc. Amer.*, **40,** 23.
SCOTT, J. F., and SINSHEIMER, R. L., 1950, *Medical Physics*, Vol. 2 (Chicago : Year Book Publishers).
SOMMER, A., and TURK, W. E., 1950, *J. Sci. Instrum.*, **27,** 113.
SPONER, H., 1942, *Rev. Mod. Phys.*, **14,** 224.
SPONER, H., and TELLER, E., 1941, *Rev. Mod. Phys.*, **13,** 76.
TANAKA, T., 1942, *Sci. Pap. Inst. Phys. Chem. Res. Tokyo*, **39,** 456.
THOMPSON, H. W., 1950, *Proc. of London Conference on Optical Instruments*, in the press.
TOUSEY, R., JOHNSON, F. S., RICHARDSON, J., and TORAN, N., 1950, *J. Opt. Soc. Amer.*, **40,** 264.
TURKEVITCH, A., and FRED, M., 1942, *Rev. Mod. Phys.*, **14,** 246.
WALSH, A. D., 1947, *Ann. Rep. Chem. Soc.*, **44,** 32; 1948, *Quart. Rev. Chem. Soc.*, **2,** 73.
WILKINS, M. H. F., and SEEDS, W. E., 1949, *Nature, Lond.*, **164,** 228.
WORLEY, R. E., 1942, *Rev. Sci. Instrum.*, **13,** 67.
WORLEY, R. E., and JENKINS, F. A., 1938, *Phys. Rev.*, **54,** 305.

ANOMALOUS FINE STRUCTURE OF HYDROGEN AND SINGLY IONIZED HELIUM

By WILLIS E. LAMB, Jr.
Department of Physics, Columbia University, New York, N.Y.

CONTENTS

	PAGE
§ 1. Introduction	20
§ 2. Theory of hydrogen-like atoms	20
2.1. Bohr–Sommerfeld theory	20
2.2. Wave mechanics	21
2.3. Dirac theory	22
§ 3. Observations of the H_α line prior to 1940	22
§ 4. Departures from the Dirac theory	24
4.1. Positron theory	25
4.2. Electromagnetic self-energy	25
4.3. Other unsuccessful attempts to account for possible discrepancies	26
4.4. Theory of finite quantum electrodynamic level shifts	27
§ 5. Possible uses of radio waves in hydrogen spectroscopy	30
5.1. Choice of method	30
5.2. Atomic beam method	31
5.3. Metastable hydrogen atoms	31
5.4. Production of a beam of metastable hydrogen atoms	33
5.5. Dissociation of molecular hydrogen	34
5.6. Excitation of hydrogen atoms by electron bombardment	35
5.7. Recoil due to bombardment	36
5.8. Detection of metastable hydrogen atoms	37
5.9. Estimate of yield	39
5.10. Radio-frequency power required	41
5.11. Breadth of resonance curves	42
5.12. Zeeman effect of the fine structure of hydrogen	42
5.13. Production of a polarized beam of atoms	44
5.14. Effect of hyperfine structure on energy levels	47
5.15. Transitions $\alpha\beta$ $2^2S_{1/2}$ ($m_s=\frac{1}{2} \rightarrow m_s=-\frac{1}{2}$)	50
§ 6. Apparatus	50
6.1. Pumps	50
6.2. Gas supply	50
6.3. Hydrogen dissociator	51
6.4. Electron bombarder	51
6.5. Quenching region	52
6.6. Detector	52
6.7. Radio-frequency techniques	52
6.8. Magnetic field	52
§ 7. Observations on hydrogen fine structure	53
7.1. Procedure	53
7.2. Resonance curves	54
7.3. Symmetry of resonances	54
7.4. Results	55
7.5. Sharp resonances	55
§ 8. Fine structure of singly ionized helium	58
8.1. Theoretical expectations for He II level shifts	58
8.2. Spectroscopic observations of He II fine structure	58
8.3. Microwave method for study of He II fine structure	59
8.4. Results for He II fine structure	61
References	62

ABSTRACT. An account is given of theoretical and spectroscopic studies of the fine structure of the $n=2$ levels of atomic hydrogen and singly ionized helium, and of the work of Lamb, Retherford and Skinner in which microwave methods are used to establish deviations from the Dirac theory. A qualitative description of the theoretical explanation in terms of quantum electrodynamic level shifts is followed by comparison with observations.

§1. INTRODUCTION

ONE of the most severe tests of any theory of the elementary particles and their laws of interaction is to be found in a comparison of predicted atomic energy levels with observation. The hydrogen-like atoms are the simplest ones in existence, consisting of a single negatively charged electron, and a positively charged nucleus. As atomic theory has developed, beginning in 1911–1913 with the Bohr–Rutherford atomic model, analyses of the spectra of hydrogen-like atoms have played a central role, either to call attention to a deficiency in existing theory, or to confirm the predictions of a newly proposed theory.

Since the close of the war in 1945, a number of observations have established discrepancies with the theory previously maintained. Fortunately much progress has already been made toward a satisfactory explanation of these phenomena, and the theory has thereby been considerably refined. The purpose of this Report is to give an account of experiments performed with the help of microwaves by Lamb and Retherford (1947, 1949a, 1950)* on the fine structure of hydrogen, and by Lamb and Skinner (1950) on the fine structure of singly ionized helium, and to describe the subsequent improvements in the theory, at least in a qualitative and intuitive manner. For this purpose, it will be convenient to review the status of the theory prior to 1947, and its agreement with spectroscopic observations.

§2. THEORY OF HYDROGEN-LIKE ATOMS

2.1. *Bohr–Sommerfeld Theory*

The energy levels of a hydrogen-like atom (for a general discussion see White 1934) are largely independent of the properties of the nucleus other than charge. The existence of a nuclear spin and magnetic moment may lead to hyperfine structure. Since the nucleus is not infinitely heavy, there are small reduced mass effects. For the present purposes, however, the nucleus may be treated as a fixed point charge Ze exerting an inverse square law Coulomb force upon the electron of mass m and charge $-e$. According to the Bohr theory of 1913, the electron could be bound to the nucleus only in certain discrete 'stationary' states of energy

$$E_n = -Z^2hcR/n^2, \qquad \ldots\ldots(1)$$

where R is the Rydberg wave number, $h=2\pi\hbar$ is Planck's constant, c is the velocity of light, and n the principal quantum number capable only of values 1, 2, 3, ... In the lowest state, $n=1$, the electron executed a circular motion of radius $a_0=\hbar^2/me^2\sim 0{\cdot}52\times10^{-8}$ cm. At the next energy levels, $n=2$, there

* A later value for the term shift of 1062 ± 5 Mc/s. is reported in the 1949 reference. The full account of this work is to appear in *The Physical Review* in four parts: I. History and preliminary observations with first apparatus. II. Qualitative studies made with new apparatus. III. Theory necessary for interpretation of refined measurements. IV. Precision observations, results and comparison with theories.

were two possible modes of motion, one circular orbit with radius $4a_0$, and one elliptic. For higher values of n, there were increasingly many modes of motion having the same energy E_n. The fact that two or more quantum states had the same energy was a special consequence of the assumption of a Coulomb attraction between electron and nucleus.

The Bohr theory gave a very precise agreement with the observations of wavelengths of many spectral lines emitted by hydrogen-like atoms. One deficiency, however, was that the Balmer radiation emitted in a transition from a higher state to the $n=2$ state actually consisted of a close doublet. This 'fine structure' had been discovered in 1887 by Michelson and Morley (1887).

In 1916 Sommerfeld (1916) found that the relativistic variation of mass with velocity could account for a dependence of the energy on the eccentricity of the orbit. For $n=2$, and small atomic number, the two levels differed by

$$\Delta E_2 = \frac{1}{16}\alpha^2 Z^4 hcR, \qquad \ldots\ldots(2)$$

where $\alpha \sim 1/137$ is the fine structure constant. For $Z=1$, ΔE_2 amounted to $0{\cdot}365\ \mathrm{cm}^{-1}$, and agreed in order of magnitude with the observed doublet separation.

2.2. *Wave Mechanics*

With the advent of non-relativistic quantum mechanics the situation was somewhat changed. The orbits given by the Bohr–Sommerfeld quantum conditions were replaced by stationary states whose statistical properties were described by wave functions, ψ_{nlm}, which were the admissible solutions of the Schrödinger wave equation. The states were described by the principal quantum number $n=1, 2, 3, \ldots$, an orbital angular momentum quantum number $l=0, 1, \ldots, n-1$, and a magnetic quantum number $m=-l, -l+1, \ldots, +l$. The energy levels for a Coulomb field agreed exactly with those of the Bohr theory and depended only on n.

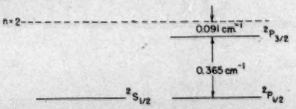


Figure 1. Fine structure of $n=2$ levels of hydrogen according to the Dirac theory. The dotted line indicates the position of the $n=2$ level according to the Bohr theory.

It was found that when wave mechanics was modified to include relativistic mass variation, the splitting of the energy levels $n=2$, $l=1$, 0 was 8/3 the value obtained by Sommerfeld. Only when effects of electron spin and magnetic moment postulated by Uhlenbeck and Goudsmit with the spin–orbit interaction deduced by Thomas were considered did the fine structure separation turn out to have the Sommerfeld value. There was then the difference that the extra degree of freedom associated with the spin doubled the number of quantum states. The energy levels in zero magnetic field are shown in Figure 1.

The $2^2S_{1/2}$ level coincided exactly in energy with the $2^2P_{1/2}$ level. The $2^2P_{3/2}$ and $2^2P_{1/2}$ states differed in energy because of spin–orbit interaction by the amount predicted by Sommerfeld's theory as given in equation (1). The selection rules and intensities for the Balmer components were worked out

according to quantum mechanics by Sommerfeld and Unsöld (1926), and the theoretical intensities were in fairly good, although not exact, agreement with the observations discussed in § 3.

The derivation of the above results was unsatisfactory because relativistic and spin corrections were added as an afterthought to the quantum mechanical equations. In fact, the calculation was really ambiguous, since the indeterminate ratio 0/0 appeared and had to be interpreted as unity in order to obtain the correct result.

2.3. *Dirac Theory*

A much more satisfactory treatment was found in 1928 by Dirac (1928) who deduced a thoroughly relativistic wave equation. A particle described by this equation was found to be automatically endowed with the desired properties of relativistic mass variation, spin, magnetic moment, and the corresponding spin–orbit interaction. Furthermore, the calculations gave an unambiguous result.

§ 3. OBSERVATIONS OF THE H_α LINE PRIOR TO 1940

An enormous amount of spectroscopic work has been done to determine exactly the extent of the agreement of observations with the prediction of the Dirac theory. The spectral lines studied included the Balmer doublets of

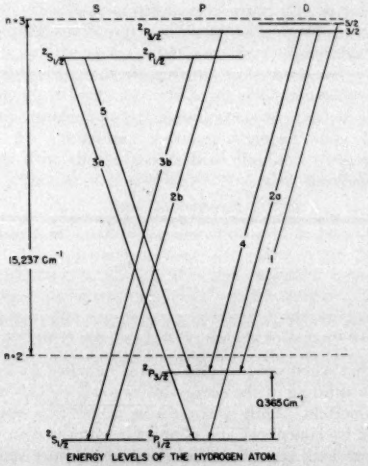


Figure 2. Energy levels of the hydrogen atom according to Dirac theory showing the component transitions of H_α line.

hydrogen (and later deuterium) and singly ionized helium. For the sake of brevity the discussion here will be limited to the H_α (and D_α) lines emitted in the transition $n=3$ to $n=2$.

The energy levels of interest in a discussion of the H_α line are shown in Figure 2, with the allowed transitions indicated by the arrows. The predicted

frequencies and intensities are shown in Figure 3. The observed H_α line should therefore consist in principle of five components. In practice, however, the individual lines are broadened by Doppler effect so that no real resolution of components 1 and 4, or even 2 and 3, can be obtained. The weak 5th component has never been clearly resolved, although an indication of its presence has lately been obtained. In earlier work the observers merely measured the separations of the unresolved complexes 2–3 and 1–4. The values obtained are shown in

Table 1. Summary of Spectroscopic Observations on H_α Line

Year	Authors	Results ($cm^{-1} \times 10^3$) Doublet separation	$\tilde{\nu}_{12}$	$\tilde{\nu}_{23}$
1887	Michelson and Morley	253		
1891	Ebert	307		
1895	Michelson	323		
1904	Houstoun	153		
1912	Buisson and Fabry	306		
1916	Paschen and Meissner	288		
1917	Merton and Nicholson	306		
1920	Merton	336		
1921	McLennan and Lowe	358		
1922	Gehrcke and Lau	293		
1922	Oldenberg	325		
1923	Geddes	339		
1923	McLennan and Shrum	332		
1924	Shrum	332		
1925	Janicki and Lau	307		
1925	Janicki and Shrum	302		
1925	Hansen	316		
1926	Houston	315		
1927	Kent, Taylor and Pearson	318		
1928	Harrison	304		
1934	Houston and Hsieh	308–312		
1934	Williams and Gibbs	304	308	
		317 (D)	321 (D)	
1934	Kopfermann		322 (D)	105 (D)
1934	Williams and Gibbs	314		
Dec.		318 (D)		
1935	Spedding, Shane and Grace		329 (D)	
1937	Houston		317–320	94–120
1937	Heyden		319 (D)	
1938	Williams		321 (D)	130 (D)
1940	Drinkwater, Richardson and Williams		316	132
			317 (D)	119 (D)
1928	Dirac's Theory		328	108

Table 1. According to the Dirac theory the separation of the centres of gravity ought to be about 0·317 cm^{-1}, assuming the correctness of the theoretical intensity ratios. Although no clear-cut discrepancy can be found in these data, there was a considerable amount of theoretical discussion as to the possible causes of any discrepancy from the theory. This is discussed in § 4 below.

Beginning with improved observations of Houston (1937), it became possible to attempt a more detailed analysis of the H_α line. Williams (1938) studied the D_α pattern and, because of the reduced Doppler broadening, was able for the

first time to resolve component 3. Both Houston and Williams found that the separations 2–3 and 2–1 departed from the values given by Dirac theory, as shown in Table 1.

Pasternack (1938) emphasized that these results could be interpreted if the $2^2S_{1/2}$ level were raised by about 0·03 cm^{-1} relative to the $2^2P_{1/2}$ level. In this case, Figure 3 should be replaced by Figure 4. An account will now be given of the many attempts at a theoretical explanation of such a shift.

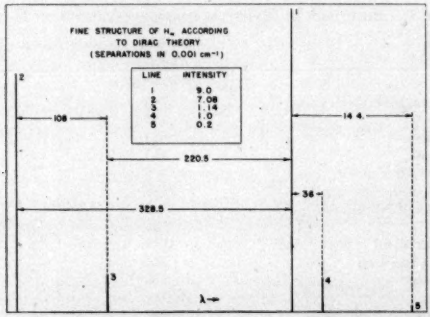


Figure 3. Fine structure pattern of H_α according to Dirac theory.

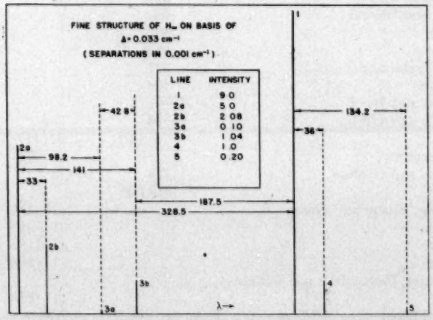


Figure 4. Fine structure pattern of H_α assuming an upward displacement of $2^2S_{1/2}$ by 0·033 cm^{-1}, and of $3^2S_{1/2}$ by $(2/3)^3$ that amount.

§ 4. DEPARTURES FROM THE DIRAC THEORY

The discrepancies between Dirac's theory and observations which have been considered include (*a*) modification of the Dirac equation or a change in its interpretation, (*b*) departure from the Coulomb law, probably at short distances from the nucleus, (*c*) effects due to the interaction of the electron with the electromagnetic field.

The Dirac equation itself led to some strange consequences. Besides the states of positive energy which appeared to agree closely with observation there

were also predicted states of negative energy. To be sure, such states also appeared in pre-quantum relativistic mechanics, but could be ignored because discontinuous changes in energy could not occur. In quantum theory, however, transitions to such states would occur rapidly. All attempts to rule out the negative energy states without spoiling the other desirable features of the theory were unsuccessful, nor was it possible simply to ignore the transitions between positive and negative energy states. Yet these transitions should have the result that all electrons should fall into negative energy states in a very short time and there exhibit properties very different from those observed.

4.1. *Positron Theory*

Dirac (1931) found a partial solution to this difficulty by supposing that *all* the negative energy states were filled with electrons. According to the Pauli exclusion principle the positive energy electrons would then be unable to fall into the negative energy states. On the basis of this theory the vacuum state was characterized by full occupancy of the negative energy states and by vacant positive energy states. There was not supposed to be any observable charge density due to the vacuum electrons. A one-electron state could be obtained by adding an electron in a positive energy state. If instead, an electron were removed from a negative energy state, the resulting hole was identified as a positive particle of positive energy. The discovery of the positron by Anderson in 1932 (Anderson 1932) lent strong support to the theory. However, there were many difficulties which had to be resolved. Even if the infinite charge density of the vacuum state were ignored, the vacuum was influenced by external electromagnetic fields, and the theory did not give finite answers to physical questions. The era of 'subtraction physics' began, in which attempts were made to subtract from the infinite results all but a finite observable residue.

Calculation of the polarization of vacuum by an electric field led to the view that any charge distribution induced an infinite charge density. All but a finite amount of this could be interpreted as proportional to the original charge density. If the constant of proportionality were finite (and small), it would be natural to consolidate this part of the induced charge with the original, and thereby 're-normalize' all charges. Owing to the divergences of positron theory, this constant of proportionality was neither finite nor small. Still, it was felt that if the theory could be made convergent by some improvements, the constant would be small. On this basis, all divergences could be attributed to the renormalization of charge, and the remaining finite polarization charge density could be taken as physically observable. In this way, Uehling (1935) calculated a correction to the Coulomb field, and found that the $2^2S_{1/2}$ level was *lowered* by $0{\cdot}9 \times 10^{-3}$ cm^{-1} (27 Mc/s.) which was far too small and in the wrong direction to explain the discrepancies which were suspected at the time. In any case, such calculations were often unsatisfactory from the point of view of relativistic and gauge invariance.

4.2. *Electromagnetic Self-energy*

According to the classical electromagnetic theory a point charge had an infinite electrostatic energy. According to relativity theory energy and mass were related to one another. In order to obtain a finite mass of the electron on a purely electromagnetic basis, it would therefore be necessary to assign the electron an extended structure. Attempts to form a satisfactory relativistic theory of an

extended charged particle failed. After the application of quantum mechanics to matter, it was applied to the problem of the electromagnetic field and its interaction with charged particles. Again the story was one of some successes and some failures. A good account was given of the phenomena of emission, absorption and scattering of radiation by matter, including spontaneous emission as postulated by Einstein in 1917. From the quantum electrodynamic point of view, spontaneous emission could be regarded as emission stimulated by the quantum zero point fluctuations in the electromagnetic fields in the vacuum or lowest energy state.

When, however, the energy of interaction of an electron with the quantized electromagnetic field was calculated by Oppenheimer (1930), an infinite result was obtained. According to his calculations the frequencies of all spectral lines should be displaced by an infinite amount from the values predicted without consideration of radiative interactions. It was pointed out by Weisskopf (1934, see also 1939) that this infinite self-energy could be ascribed to the zero point fluctuations in the vacuum electromagnetic field. These fluctuations were needed for spontaneous emission and yet led to violent disagreement with observations in other respects. The theory of the electromagnetic self-energy remained in this highly unsatisfactory state until 1947.

4.3. *Other Unsuccessful Attempts to Account for Possible Discrepancies*

As early as 1933 Kemble and Present (1933) recognized that a shift upward of the $2^2S_{1/2}$ level might account for the results of Houston and Hsieh (1934), and considered the effect of a short range modification in the Coulomb law. The necessary departures were so large as to seem unreasonable. Other authors have discussed the effect of postulated modifications of the Coulomb field, among them Jauch (1940), Sommerfeld (1941 a, b), Caldirola (1948) and David (1949). The Uehling correction mentioned above was far too small and in the wrong direction. The departure from the Coulomb law arising from the finite size of the deuteron gave an elevation of $2^2S_{1/2}$, but by an even smaller amount.

In 1935, Bechert and Meixner (1935) showed that reduced mass effects did not remove the $2^2S_{1/2}$–$2^2P_{1/2}$ degeneracy, and that hyperfine structure could not account for any discrepancy (Bechert and Meixner 1935, Breit and Brown 1948).

According to meson theories of nuclear forces a proton can dissociate into a neutron and a positive meson. As a result, a proton spends a certain fraction of the time in the dissociated form, and there is on the average a charge density of positive mesons about the central nucleon. Since the radius of this meson cloud is of the order of magnitude of the range of nuclear forces, one would not intuitively expect the resulting departures from the Coulomb field to be even as large as those for the deuteron. Nevertheless, in 1939 Fröhlich, Heitler and Kahn (1939) calculated the $2^2S_{1/2}$ level displacement to be expected according to a meson theory and obtained an upward shift of about 0·03 cm^{-1} in agreement with Pasternack's requirement. A closer examination (Lamb 1939, 1940, Fröhlich, Heitler and Kahn 1939, Kahn 1941, Blatt 1945, Slotnick and Heitler 1949) of these calculations revealed that although the integrals which occurred were convergent, there was a hidden divergence which invalidated the result. In effect, the proton was required to be dissociated more than 100% of the time.

While such matters were being debated, Drinkwater, Richardson and Williams (1940) published the results of very careful spectroscopic observations,

Although their results were similar to those of Houston and Williams, they possibly over-emphasized the importance of intensity anomalies and impurities in the discharge source and stated "We have come to the conclusion that no real evidence has yet been found to show that the fine structure patterns of H_α and D_α depart substantially from the values calculated on the basis of Dirac's theory". This verdict was accepted by many of the theorists who had been attempting an explanation of such discrepancies.

Post-war measurements by Giulotto (1947a, b) and by Kuhn and Series (1948)* have again pointed the way to a clear-cut departure from the theory in agreement with the analyses of Houston and Williams.

The H_α complex has two main peaks separated by about 0·317 cm^{-1} and, in the work of Williams with deuterium, the lines had Doppler widths of 0·120 cm^{-1}. Under these circumstances a really precise spectroscopic check of the theory is exceedingly difficult. It is made even more so by the fact that the intensities of the components often depart from the theoretically predicted values and vary with the discharge conditions.

Only the use of an atomic beam source to reduce Doppler effect would seem to offer much hope for a real resolution of the H_α complex. The radiative width of the sharpest line would be only about 0·001 cm^{-1} (30 Mc/s.) but, according to estimates made by Mack and Barkofsky (1942), the attainable width would be about ten times larger.

4.4. *Theory of Finite Quantum Electrodynamic Level Shifts*

After the microwave confirmation in 1947 (Lamb and Retherford 1947) of a level shift of the order of magnitude 1,000 Mc/s., Bethe (1947) found that the quantum electrodynamics really had hidden behind its divergences a physical content which was in very close agreement with the observations. The crucial idea due to Bethe was re-normalization of mass.† The quantum theory of radiation predicts that a free electron should have an infinite mass. Some explanation must therefore be found for the fact that the observed mass is finite, but Bethe, recognizing that this was outside the scope of the theory for the moment, proposed to ignore the electromagnetic mass completely. For an electron bound in a hydrogen atom, an infinite energy occurs, but this is merely a manifestation of the infinite electromagnetic mass which should be eliminated in some future theory. If the mass terms are properly subtracted a finite remainder is obtained which would be zero for a free electron. In the case of a bound electron the force field modifies the effect of the electromagnetic field and a finite displacement of the energy levels results.

A qualitative explanation of this was given by Welton (1948). He considered the quantized electric field in vacuum as described by a Fourier expansion in terms of plane waves

$$\mathbf{E}_\mathbf{k} \exp\{i\mathbf{k}.\mathbf{r} - i\omega t\}, \qquad \ldots\ldots(3)$$

where $\mathbf{E}_\mathbf{k}$ is the amplitude of the wave with propagation vector $\mathbf{k}$ and circular frequency $\omega = kc$. When the radiation field is quantized it is found that the

* These workers have lately obtained an improved value of 0·0369±0·0016 cm^{-1} (*Proc. Roy. Soc.* A, 1950, **202**, 127).

† Actually, Serpe and Kramers used this notion earlier (*Physica*, 1940, **7**, 133) to treat a related problem. The mass renormalization terms were not so easily recognized by earlier workers as were the corresponding charge terms, either because their calculations were too divergent, not sufficiently explicit, or else contained errors.

lowest energy state is not one with zero electromagnetic fields, but there are zero point oscillations, similar to those for the simple harmonic oscillator treated according to quantum mechanics. The zero point energy associated with each Fourier component is just $\frac{1}{2}\hbar\omega$ so that the amplitude $\mathbf{E}_\mathbf{k}$ is determined by the equation

$$\frac{1}{8\pi}(\mathbf{E}^2+\mathbf{H}^2)=\frac{1}{4\pi}\int d\mathbf{k}|\,\mathbf{E}_\mathbf{k}|^2=\int\frac{\hbar\omega}{2}\,dN(\mathbf{k}), \qquad \ldots\ldots(4)$$

where

$$dN(\mathbf{k})=2\,d\mathbf{k}/(2\pi)^3 \qquad \ldots\ldots(5)$$

is the number of radiation oscillators whose wave vectors lie in $d\mathbf{k}$. Hence

$$|\,\mathbf{E}_\mathbf{k}|^2=\hbar\omega/2\pi^2. \qquad \ldots\ldots(6)$$

These fluctuating electric fields, acting on an electron, cause it to have a rapidly variable position. In effect, the point electron becomes a sphere of a certain radius. Such an electron in a hydrogen atom is not so strongly attracted to the nucleus at short distances. As a result, states of zero orbital angular momentum are raised in energy relative to other states in which the electron has a small probability of being found near the nucleus.

The change in energy may be estimated as follows. The position of the electron is written as $\mathbf{r}+\boldsymbol{\delta}\mathbf{r}$, where $\boldsymbol{\delta}\mathbf{r}$ is the change in position due to the zero point field oscillations. The average value of $\boldsymbol{\delta}\mathbf{r}$ is zero, but $(\boldsymbol{\delta}\mathbf{r})^2$ will have a non-zero average value $(\boldsymbol{\delta}\mathbf{r})^2_{\rm av}$. The kinetic energy of these oscillations contributes to the electromagnetic mass of the electron and can be ignored. For a bound electron there is also an effect on the potential energy, $V(\mathbf{r})$, which becomes $V(\mathbf{r}+\boldsymbol{\delta}\mathbf{r})$. Since $\boldsymbol{\delta}\mathbf{r}$ is small, this may be expanded in a Taylor's series giving

$$V(\mathbf{r}+\boldsymbol{\delta}\mathbf{r})=V(\mathbf{r})+(\boldsymbol{\delta}\mathbf{r}\cdot\nabla)V(\mathbf{r})+\tfrac{1}{2}(\boldsymbol{\delta}\mathbf{r}\cdot\nabla)^2V(\mathbf{r}) \qquad \ldots\ldots(7)$$

with neglect of higher terms.

In taking a time average the terms linear in $\boldsymbol{\delta}\mathbf{r}$ drop out, so that one finds a change in the potential energy of $\delta V=\frac{1}{6}(\boldsymbol{\delta}\mathbf{r})^2_{\rm av}\nabla^2V$ due to the zero point field oscillations.

To calculate $(\boldsymbol{\delta}\mathbf{r})^2_{\rm av}$ use is made of the equation of motion

$$m\boldsymbol{\delta}\ddot{\mathbf{r}}=e\mathbf{E}. \qquad \ldots\ldots(8)$$

The electric field $\mathbf{E}$ is a sum of terms of the form of equation (3). For the kth component of the forced oscillation

$$m\boldsymbol{\delta}\mathbf{r}_\mathbf{k}=-\frac{e\mathbf{E}_\mathbf{k}}{\omega_\mathbf{k}^2}, \qquad \ldots\ldots(9)$$

and for the resultant squared displacement

$$(\boldsymbol{\delta}\mathbf{r})^2_{\rm av}=\int d\mathbf{k}(\boldsymbol{\delta}\mathbf{r}_\mathbf{k})^2_{\rm av}=\frac{e^2}{m^2}\int d\mathbf{k}\frac{|\,\mathbf{E}_\mathbf{k}|^2}{\omega_\mathbf{k}{}^4}=\frac{2}{\pi}\frac{e^2}{\hbar c}\left(\frac{\hbar}{mc}\right)^2\int_{k_1}^{k_2}\frac{dk}{k}, \qquad \ldots\ldots(10)$$

where the limits of integration k_1 and k_2 have to be inserted to keep the integral from diverging logarithmically. The effect of the fluctuations on the potential energy is then

$$\delta V=\frac{1}{3\pi}\frac{e^2}{\hbar c}\left(\frac{\hbar}{mc}\right)^2\nabla^2V\log\frac{k_2}{k_1}. \qquad \ldots\ldots(11)$$

When this is averaged for the quantum mechanical state ψ, the resulting energy shift is

$$\delta E=\frac{4}{3}\alpha Ze^2\left(\frac{\hbar}{mc}\right)^2|\psi(0)|^2\log\frac{k_2}{k_1} \qquad \ldots\ldots(12)$$

so that only s states are shifted. The equation obtained by Bethe from non-relativistic quantum electrodynamics was just of this form, except that the limit of integration k_2 was determined by the calculation, and a physically reasonable value for k_2 could be guessed. The lower limit k_1 enters the result because the bound electron can interact effectively with radiation oscillators which have an energy large compared with the excitation energy of the atom. Bethe found in his calculation that

$$hck_1 \simeq 17{\cdot}8hcR. \qquad \ldots\ldots(13)$$

The upper limit k_2 was taken to be mc/h, because of the expectation that a relativistic calculation would lead to convergence for light quantum energies above the rest energy of an electron. The logarithm appearing in equation (12) was accordingly

$$L = \log \frac{k_1}{k_2} \sim \log \frac{mc^2}{17{\cdot}8hcR} \sim 7{\cdot}63. \qquad \ldots\ldots(14)$$

The predicted shift of the $2^2S_{1/2}$ level of hydrogen, expressed in frequency units is then $\delta E \sim 1{,}040$ Mc/s. This result was in very good agreement with the first rough microwave measurements.

A relativistic generalization of Bethe's calculation was made by Kroll and Lamb (1949), and by French and Weisskopf (1949). Their result may be written

$$\delta E(m, l) = \frac{8Z^4}{n^3} \frac{\alpha^3 hcR}{3\pi} L, \qquad \ldots\ldots(15)$$

where now

$$\left.\begin{aligned} L &= \log \frac{k_2}{k_1(n,0)} - \log 2 + \frac{5}{6} - \frac{1}{5} \quad \text{for} \quad l=0, \\ L &= \log \frac{R}{k_1(n,l)} + \frac{3}{8} \frac{C(l,j)}{2l+1} \quad \text{for} \quad l \neq 0, \end{aligned}\right\} \qquad \ldots\ldots(16)$$

where

$$C(j,l) = \begin{cases} 1/(l+1) & \text{for} \quad j = l + \frac{1}{2} \\ -1/l & \text{for} \quad j = l - \frac{1}{2} \end{cases}$$

for the displacement of the state n, l, $j = l \pm \frac{1}{2}$. For hydrogen, the necessary constants $k_1(n, l)$ have been calculated more accurately by Bethe, Brown and Stehn (1950), and the elevation of $2^2S_{1/2}$ relative to $2^2P_{1/2}$ should be 1,051·41 Mc/s. Equation (15) includes the Uehling (1935) vacuum polarization effects, as well as contributions from the spin–orbit interaction due to the anomalous magnetic moment of the electron (Breit 1947, Kusch and Foley 1948, Schwinger 1948).

The above calculations were performed with the older formulation of relativistic electrodynamics according to which the subtraction of the infinite energies associated with renormalization of mass and charge in a covariant fashion is rather difficult. Subsequently, the same result has also been obtained by a thoroughly relativistic treatment of the problem according to covariant methods developed independently by Tomonaga (1946 ff.), Schwinger (1948, 1949 a, b), Feynman (1949) and Dyson (1949). It should be pointed out, however, that all of these calculations are incomplete in at least two well-defined respects. They were performed by perturbation theory only to second order in the interaction between the electron and the radiation field. Also, certain effects associated with the binding of the electron were neglected. It is not believed that either of these errors would account for more than a few megacycles per

second in the above shift for hydrogen, but calculations must be made to be sure of this. The possibility remains that new phenomena may still be brought to light by accurate studies of the hydrogen fine structure.

§5. POSSIBLE USES OF RADIO WAVES IN HYDROGEN SPECTROSCOPY

It was pointed out as early as 1928 by Grotrian in his book *Graphische Darstellung der Spektren von Atomen*, that the selection rules permitted optical transitions with zero change of principal quantum number n and that it should be possible with radio waves to induce such transitions among the excited states of the hydrogen atom. For $n=2$, the doublet separation $\Delta\tilde{\nu}=0{\cdot}365\ \mathrm{cm}^{-1}$ corresponds to a wavelength 2·74 cm. or a frequency of 10,949 Mc/s.

There are some German papers between 1932 and 1935 dealing with attempts to detect such a transition. At that time only spark gap oscillators of exceedingly low power output were available for such wavelengths. One had to work with a continuous spectrum of radiation, using interference methods to select a reasonably monochromatic range of wavelengths. The radiation was passed through an absorption vessel containing the excited atoms produced in a hydrogen discharge of the Wood's (1921) type, and the attenuation was determined as a function of wavelength.

The first worker, Betz (1932)* reported an absorption in the expected wavelength range of 3 cm. He also obtained selective absorptions at 9 and 27 cm. corresponding to transitions between the fine structure states of the levels with $n=3$. Three years later, working under the same professor in the same laboratory, Haase (1935) repeated the experiment more carefully and failed to find any absorptions at 9 and 27 cm. He did not study the 3 cm. region. Haase made estimates of the selective energy absorption to be expected from excited hydrogen atoms and concluded that it was too small to detect. Strangely enough, Haase gave no specific reason for the positive results obtained by Betz.

5.1. *Choice of Method*

After reading of the Betz–Haase work, Lamb and Retherford (1947) inquired whether microwave techniques, which advanced so much during the war, would now permit a successful and clear cut determination of the hydrogen fine structure by absorption of the appropriate radiation in a Wood's discharge. In the course of such considerations they came upon the idea of a different method which was then actually used. While they did not reach a definite conclusion on the possibility of the use of a Wood's tube method, the prospect was rather discouraging. In view of the extreme crudeness of the numerical estimates, however, it is by no means certain that one could not now detect transitions between the $n=2$ states of atomic hydrogen if the discharge conditions were properly chosen. From consideration of the perturbing effects of electric fields and ion collisions on the energy levels of the atom, however, it seems that even if the transitions could be detected, the results would be of greater value for an indication of the conditions within a Wood's discharge tube than for a determination of fundamental properties of an isolated hydrogen atom.

* The dispersion of radio waves in a hydrogen discharge was studied by E. E. Schneider (1934 Dissertation, Freiburg).

There are two main methods in radio-frequency spectroscopy. In one, some matter absorbs or otherwise affects the radiation passing through it. In the other, the radiation causes observable changes to take place in the matter. The former method is used in the conventional form of microwave spectroscopy (Gordy 1948, review article) in which the radiation is transmitted through a long path of absorbing gas. The latter method is used in the molecular beam radio-frequency resonance (Kellogg and Millman 1946, review article) method when the particles absorbing radiation are deflected out of the beam by an inhomogeneous magnetic field. Since the first method did not seem too promising for study of the hydrogen fine structure, Lamb and Retherford turned to an examination of the possibilities in the second method.

5.2. *Atomic Beam Method*

In the case of atomic hydrogen the 2p states decay to the 1s state with the emission of a photon of wavelength 1,216 A. in $1{\cdot}595 \times 10^{-9}$ sec., and the atom could move only about $1{\cdot}3 \times 10^{-3}$ cm. in that time, assuming a speed of 8×10^{5} cm/sec. On the other hand, the possibility exists that the $2^2S_{1/2}$ state is sufficiently metastable so that a beam of particles in this state could be used. If a transition were then induced by radio-frequency radiation or otherwise to a 2p state, a decay to 1s would take place so quickly that the number of excited atoms in the beam

Figure 5. Schematic block diagram of apparatus.

would be reduced. If one could then find a detector which responded selectively to the excited hydrogen atoms, one would have the possibility of measuring the energy difference between the metastable $2^2S_{1/2}$ state and the various 2p states. The proposed experiment might then be represented in the block diagram in Figure 5.

For this programme to be carried out, it was necessary to solve the following problems: (1) production of a beam of atoms in the $2^2S_{1/2}$ state, (2) detection of such atoms. The programme clearly depended on a knowledge of the properties of metastable hydrogen atoms. Such atoms had never been isolated in any clear-cut experiments, although they had been the subject of much speculation. The next section is devoted to the properties of metastable hydrogen atoms.

5.3. *Metastable Hydrogen Atoms*

In 1924 Compton and Russell (1924), in trying to explain the large intensity of stellar Balmer lines, pointed out that the lower of the two $n=2$ states obtained on the basis of Sommerfeld's theory should be metastable. In 1926, after the spinning electron theory was introduced, Sommerfeld and Unsöld (1926) stated that the $2^2S_{1/2}$ state should still be metastable. On the other hand, Franck and Jordan (1926, see also de Groot and Penning 1933) gave reasons for doubting that the $2^2S_{1/2}$ level would be metastable.

A large amount of experimental work by Snoek, von Keussler and others (see Bethe 1933, pp. 452–462) was carried out to test the Sommerfeld–Unsöld ideas on this point. They studied the relative intensities of the two main groups

of components of the H_α line in absorption. If the $2^2S_{1/2}$ state in the absorbing discharge were appreciably metastable, the intensity ratio of the two peaks would be affected. From their results they concluded that the $2^2S_{1/2}$ level was not appreciably metastable.

Rojansky and Van Vleck (1928) and Rojansky (1929) pointed out that the effect of an external electric field, however small, would mix the degenerate states $2^2S_{1/2}$ and $2^2P_{1/2}$ giving stationary state wave functions

$$\psi = \frac{1}{\sqrt{2}}[\psi(2^2S_{1/2}) \pm \psi(2^2P_{1/2})], \qquad \ldots\ldots(17)$$

both of which would decay at the rate $(2\tau_p)^{-1}$, so that there could be no metastability. This conclusion, although correct in a sense, is not applicable to experimentally realized situations where no stationary state exists. This was pointed out by Bethe (1933, pp. 452–455) who, in his *Handbuch* article, gave a systematic theory of the effect of an electric field on the fine structure of hydrogen. In the complete absence of a perturbing electric field, an atom in the $2^2S_{1/2}$ state should have a very long life. (Bethe estimated that the life of an isolated atom would then be several months if relativistic effects were taken into account.) In the presence of a large perturbing electric field, the result would be that of Rojansky and Van Vleck. As the electric field is increased from zero, the life-time would undergo a change from one limiting value to the other. Bethe showed that the results of von Keussler and Snoek were understandable in terms of the excitation cross sections of the various levels and the quenching of the metastability by the perturbing electric fields of the discharge. For an effective electric field of 10 volts/cm., he found that the life of the $2^2S_{1/2}$ state would be about five times the life τ_p. Over a large range of fields the life varies inversely as the square of the field strength. Under typical Wood's tube conditions Bethe estimated that the life would be about $2\tau_p$.

Breit and Teller (1940)* refined some of Bethe's calculations and discussed the metastability in connection with possible astrophysical applications. They showed that the decay mechanism of double quantum emission to the ground state was more important than the relativistic effects, and that the life of an isolated metastable atom on this basis would be one-seventh of a second.

It seemed clear from the foregoing that the $2^2S_{1/2}$ state would be markedly metastable only if the perturbing fields could be sufficiently reduced. For a beam length of 6 cm. and a speed of 8×10^5 cm/sec., a life of the order $0{\cdot}75 \times 10^{-5}$ sec. would be necessary for 37% of the atoms to survive. According to Bethe's calculations this would require a perturbing field of $\frac{1}{3}$volt/cm. or less. It would not be easy to keep the electrons and ions formed in the excitation process away from the detector with such fields unless an extremely long beam were used.

In actual fact, of course, it is now known that the $2^2S_{1/2}$ and $2^2P_{1/2}$ levels are not degenerate. This adds considerably to the stability of $2^2S_{1/2}$ against Stark effect quenching. According to a simple extension of Bethe's calculation the life-time of $2^2S_{1/2}$ in moderate electric fields becomes

$$\tau_s = \tau_p\{\hbar^2(\omega^2 + \tfrac{1}{4}\gamma^2)/|V|^2\}, \qquad \ldots\ldots(18)$$

where V is the matrix element of the perturbing electric energy and $\hbar\omega$ is the separation of the competing levels $2^2S_{1/2}$ and $2^2P_{1/2}$. (Breit and Teller already

* A corresponding theoretical analysis of double quantum emission has been given by J. A. Wheeler, *J. Opt. Soc. Amer.*, 1947, **37**, 813.

discussed the stabilizing effect of the hyperfine structure splitting and found an increase in τ_s by a factor of four. Since the actual value of $\hbar\omega$ is much larger than the hyperfine structure splitting, the stability will not be much affected by the hyperfine structure, and it will therefore be neglected here.) According to equation (18), for $\omega/2\pi = 1{,}051$ Mc/s., the life τ_s would be about 442 times longer than if there were no removal of the degeneracy.†

At the time, however, Lamb and Retherford did not take the possibility of a natural removal of the degeneracy by such a large amount very seriously, and planned to increase the stability of $2^2S_{1/2}$ by the application of a magnetic field to give a large Zeeman splitting of the states.

The presence of a magnetic field at right angles to the atomic beam would also serve to keep charged particles away from the detector. A third, and really the most important, function of the magnetic field will be discussed in § 5.11.

5.4. *Production of a Beam of Metastable Hydrogen Atoms*

A number of possible methods for the production of a beam of hydrogen atoms in the metastable $2^2S_{1/2}$ state were considered. The simplest source would be a hydrogen discharge tube with a small opening into a vacuum for the emergence of a beam. In the discharge one would have a mixture of molecular and atomic hydrogen, electrons and ions, and a small proportion of excited atoms, together with a high intensity of Lyman and Balmer radiation. The question would be whether any appreciable number of metastables could escape through the hole before being quenched by the Holtsmark field due to ions and electrons simultaneously escaping. Even with the stabilizing effect of a magnetic field, it did not seem likely. Also, there would be a very high background intensity of ultra-violet radiation as the detectors of metastables are also photosensitive. Subsequent to the success of the experiment by another method such a Wood's tube source was briefly tried, but positive results were not obtained.

The second method considered was the bombardment of molecular hydrogen with electrons in a field-free region. A possible process is the one

$$H_2 + e \rightarrow H + H^* + e', \qquad \ldots\ldots (19)$$

where H^* represents a metastable atom. The potential energy curves for the various excited electronic states of H_2 shown in Figure 6 are based on calculations by Hylleraas (1931), Present (1935) and Coolidge, James and Present (1938). The small segment of the ground state $^1\Sigma_g$ shown represents the limits of zero point vibration in the lowest vibrational state. According to the Franck–Condon principle, electron bombardment should most strongly excite the $^3\Sigma_g$ state at energies below its dissociation limit into $H + H^*$. One would, however, expect a small yield of metastable atoms for electrons of about 15 ev. Such a 'violation' of the Franck–Condon principle would be analogous to the dissociation with ionization

$$H_2 + e \rightarrow H + H^+ + e' + e'' \qquad \ldots\ldots (20)$$

observed by Bleakney (1930) at 18 volts.‡ Besides the $^3\Sigma_u$ curve shown, there are repulsive states leading to $H + H^*$, and hence fast metastable atoms should be produced at higher bombarding energies.

† This suggests the need for a re-interpretation of the experiments of Snoek, v. Keussler and others. Some of the features which need consideration are indicated in Appendix I of the paper cited in footnote on page 20.

‡ H. F. Newhall, 1942, *Phys. Rev.*, **62**, 11, has questioned this interpretation, but his apparatus may have discriminated against slow protons.

In the apparatus to be described below, Lamb and Retherford (1949b) have observed the production of metastable hydrogen atoms by this process. One difficulty with this method of formation is that the atomic fragments H + H* move off in directions oriented at random with respect to the electron beam, and hence the detector can conveniently intercept only a small fraction of them. The background effects due to the ultra-violet photons which are produced in molecular hydrogen beginning at 11·5 volts are therefore very troublesome. It was found that a third method gave a better yield.

This method required the separate production of atomic hydrogen atoms in the ground state and the subsequent bombardment of a beam of them by electrons of energy somewhat over the threshold of 10·2 volts for excitation to $2^2S_{1/2}$. The proposed experiment might now be represented by the block diagram shown in Figure 7.

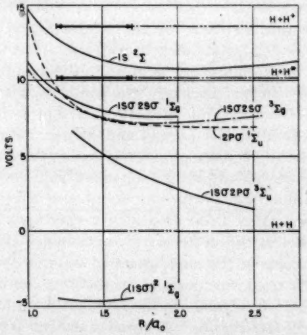


Figure 6. Electronic energy levels of the hydrogen molecule as a function of internuclear distance.

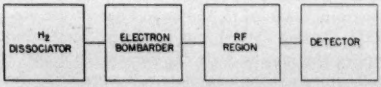


Figure 7. Modified schematic block diagram of apparatus.

5.5. *Dissociation of Molecular Hydrogen*

A number of methods are available for the production of a beam of atomic hydrogen: (i) Wood's tube, (ii) microwave discharge, (iii) thermal dissociation in a tungsten furnace. The first method is known to be very efficient, but it was felt that the ultra-violet background radiation might be troublesome. The microwave discharge method (Davis *et al.* 1949) was ruled out because the leakage of a small amount of radio-frequency power would complicate the spectroscopic observations. The last method was used in 1923 by Olmstead and Compton (1923) who measured the critical potentials of atomic hydrogen formed in a tungsten oven, described by Duffendack (1922) and Compton (1922).

Assuming thermal equilibrium, the extent of the dissociation of molecular hydrogen

$$H_2 \rightarrow 2H \qquad \ldots\ldots(21)$$

is given by

$$[P(H)]^2/P(H_2) = K(T), \qquad \ldots\ldots(22)$$

where the equilibrium constant $K(T)$ is given in atmospheres as a function of absolute temperature T by the equation

$$\log K = -\frac{21200}{T} + 1{\cdot}765 \log T - 9{\cdot}85 \times 10^{-5} T - 0{\cdot}265 \qquad \ldots\ldots(23)$$

calculated by Bonhoeffer (1927, see also Wooley, Scott and Brickwedde 1948) from data of Langmuir. Writing

$$P(H) = XP, \qquad \ldots\ldots(24)$$

where P is the total pressure

$$P = P(H) + P(H_2), \qquad \ldots\ldots(25)$$

the fractional dissociation X is determined by the equation

$$\frac{X^2}{1-X} = \frac{K(T)}{P}. \qquad \ldots\ldots(26)$$

In Figure 8 X is shown as a function of T for several pressures.

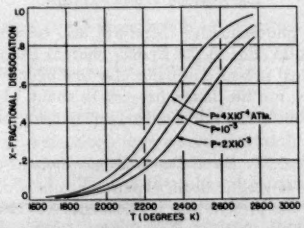


Figure 8. Thermal dissociation of molecular hydrogen.

No direct measurement of pressure in the tungsten oven could be conveniently made, but from the observed variation of yield with temperature when the experiment was successful, it seemed reasonable that it was about 10^{-3} atmosphere in a typical case, and that the dissociation was 64% complete at $T = 2{,}500°$ K. At this temperature the most probable velocity in the beam was 8×10^5 cm/sec.

5.6. *Excitation of Hydrogen Atoms by Electron Bombardment*

The cross sections for various electron excitation processes in hydrogen have been calculated by Bethe (1933, p. 507) according to the Born approximation with neglect of electron exchange. These cross sections are plotted as functions of electron energy in Figure 9. The maximum cross section for 2s ($\sigma_{max} = 2{\cdot}2 \times 10^{-17}$ cm^2) is about one-ninth that for 2p, while that for 3p is smaller than that for 2s by a factor of 11. As a result, the yield of 2s by cascade from 3p may be neglected. The maximum yield for 2s comes at 14·8 volts, but does not fall below half the maximum value down to 11 volts.

Besides the errors inherent in the Born approximation, the neglect of electron exchange may be particularly serious for excitation of the optically inaccessible state 2s, and one may hope to obtain a much larger cross section with a maximum nearer the threshold if exchange is included. Since no reliable calculations of exchange excitation, even with the Born approximation, seemed to exist for hydrogen, Lamb and Retherford used the lower cross section $\sigma = 10^{-17}$ cm^2, hoping to be on the safe side.

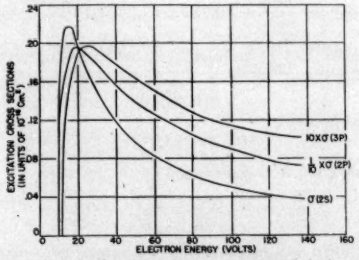


Figure 9. Electronic excitation cross sections for states of atomic hydrogen calculated by Born approximation without exchange.

If the detector is photosensitive there will also be a signal with the same threshold energy of 10·2v. due to the Lyman photons from atoms excited to the 2p states. Any residual pressure of molecular hydrogen will also give photons. The threshold energy for the latter process is about 11·5 ev. Only a small fraction of these photons can reach the detector, but any background effects are undesirable.

5.7. *Recoil Due to Bombardment*

In the usual form of atomic beam experiment one works with a very well collimated beam. In the present case it is necessary to bombard the hydrogen atoms after they leave the source, since Stark quenching must be minimized. One must then choose between bombardment at right angles to the beam, along it, or against it. The last two methods would not destroy the unidirectionality of the beam of atoms if the bombardment was at a voltage just above the threshold. Such a choice would be natural for other atoms but not for metastable hydrogen atoms. In this case, one would be forced to have the magnetic field parallel to the electron beam. If the electrons were sent against the atomic beam the metastable atoms would subsequently pass through an electric field which would quench them. If the electrons were sent along the beam they could not easily be kept from the detector. Methods based on difference in time of flight were rejected for the initial experiments on intensity grounds.

The first choice necessarily leads to a transverse recoil, and, what is worse, an indefinite one, so that an originally well collimated beam is diffused by the bombardment. This means that the use of slits of the order of 0·001 in. width as in the usual atomic beam work is ruled out. On the other hand, one does not have the problem met there of detecting the small deflections of the beam in inhomogeneous magnetic fields, since the effect of the radio-frequency radiation is much more easily distinguished here by the removal of the excitation of the absorbing atom.

The distribution of recoil angles is easily calculated and the results for typical cases are shown in Figure 10. Even at the threshold there is a distribution in horizontal recoil angle ranging (with larger than half the maximum probability)

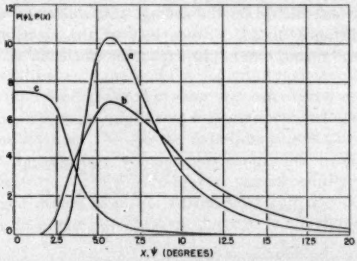


Figure 10. Distributions in recoil angles. Curve (*a*): distribution in horizontal recoil angles ψ experienced by a hydrogen atom excited to the 2s state by electrons of energy just above the threshold. The temperature of the source is taken to be 2,600° K. The total probability would be unity if the angle were measured in radians. Curves (*b*) and (*c*): distributions in horizontal recoil angles ψ and vertical recoil angles χ for hydrogen atoms excited to 2s by electrons of energy 13·6 ev.

from 4° to 8·7° because of the distribution of velocities in the atomic stream. At 13·6 ev. the distribution broadens to range from 3·6° to 10·8°, and there is a distribution of vertical recoil angles between $\pm 3{\cdot}4°$.

5.8. *Detection of Metastable Hydrogen Atoms*

The atoms excited by the electron bombardment to the metastable $2^2S_{1/2}$ state are supposed to move off to a suitable detector. Several possible methods of detection were considered. Two of these had previously been used to detect metastable atoms other than hydrogen. It was discovered by Webb (1924) that metastable atoms of mercury could eject electrons from metals. Later work by Oliphant (1929) showed that the same was true of helium metastable atoms. A theory of this process was given by Massey (1930, 1931) and by Cobas and Lamb (1944). If the excited atom comes close to the surface of a metal, it may be energetically possible for a collision of the second kind to occur in which the atom returns to the ground state and an electron is liberated from the system to take up the excess energy. One possibility for such a process is shown in Figure 11. The energetic condition for this is that the excitation energy of the atom should exceed the work function of the metal. In the case of helium the excitation energy is about 20 ev. and the condition should be satisfied for all metals. In the case of mercury the excitation energy of the 2^3P_0 state is only 4·68 volts, and the condition becomes much more critically dependent on the work function of the surface. In fact, Sonkin (1933) showed that the efficiency of detection of mercury metastables was very sensitive, by a factor of 100 or more to the presence of surface contaminations. On the other hand, Dorrestein (1942) found that the efficiency of detection of helium metastables on outgassed platinum was about 40% and was subject only to relatively small fluctuations over a day's run. No information was available on the detection of hydrogen metastables. In this case, however, the excitation energy, 10·2 volts, is well above

the work function of metals, and one would at first expect a high efficiency of detection as with helium.

According to rough calculations (Sternberg, private communication) of the type mentioned above, metastable hydrogen atoms of thermal energy moving in toward a metal surface should on the average cause electron ejection before the atom reaches a distance of 2 A. Since half of the electrons emitted would probably go into the metal, one might expect an efficiency of the order of 50%.

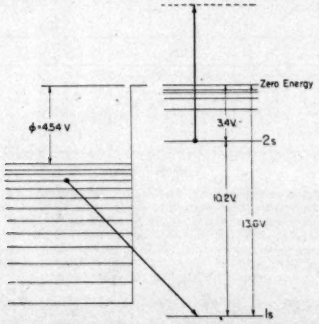


Figure 11. A possible mechanism for ejection of electrons from tungsten by metastable hydrogen atoms due to the Coulomb interaction between atomic and metallic electrons.

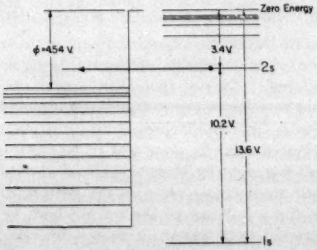


Figure 12. Auto-ionization of a metastable hydrogen atom in the vicinity of a tungsten surface.

Nevertheless, Lamb and Retherford were not very confident about the possibility of detecting hydrogen metastables by the above mechanism. These doubts arose from their consideration of another method based on the observation by Buehl (1933) that metastable atoms of mercury falling on hot molybdenum are re-emitted as ions. This process is essentially the surface ionization detection method used for molecular beams where an alkali atom falls on a hot tungsten surface, an electron is captured by the metal, and the resulting ion is evaporated from the surface. Such a process might be represented as shown in Figure 12 and may take place if the energy inequality

$$\phi > I \qquad \text{.....(27)}$$

holds.

In the case of hydrogen atoms in the 2s state, the ionization potential $I=3{\cdot}4$ volts. Most surfaces which can be maintained under the conditions of the experiment would probably have a work function $\phi>3{\cdot}4$ volts, and capture of the 2s electron by the surface is energetically possible. A rough estimate (Sternberg, private communication) of the mean distance at which the process would occur gives 5 A., i.e. a larger distance than that estimated for electron ejection. If the resulting proton could escape from the surface with appreciable probability, one would have a method for detecting metastable hydrogen atoms. If, however, the proton is neutralized in some way not involving electron emission, the process of electron capture might very well compete seriously with the desired process of electron ejection. When the experiment was successful, Lamb and Retherford found that electron ejection did occur, but probably not with the expected high efficiency. They have looked unsuccessfully for re-emission of metastable hydrogen atoms as protons from cold and non-outgassed tungsten. Further investigation of the detection mechanism is clearly desirable.

5.9. *Estimate of Yield*

Before undertaking their experiment, Lamb and Retherford made a rough estimate of the yield of metastable hydrogen atoms and the resulting electron currents to be expected under typical experimental conditions. Consider the atoms emerging from the tungsten oven. According to ideal kinetic gas theory the number of hydrogen atoms escaping per second at an angle θ with the normal to the wall into a solid angle range Ω through a small opening of area a is

$$\text{Rate of escape} = \frac{n_0 a v \Omega X \cos\theta}{4\pi}, \qquad \ldots\ldots(28)$$

where n_0 is the number of atoms and molecules per unit volume, X the fractional dissociation, and v is an average velocity of the atoms in the oven.

Few of these atoms can reach the detector plate directly because it is off-set to allow for the small recoil angle. If, however, it were not so displaced, and no obstacles were in the way, a fraction of the atoms emerging from the oven slit would reach the detector of area A at a distance R from the slit. For this, $\Omega = A/R^2$ and $\cos\theta \sim 1$. Let f be the fraction of atoms which are excited to the 2s state as they pass through the electron gun region. The excited atoms experience a deflection approximately sufficient to allow them to strike the displaced detector target. As there is a spread of recoil angles, the metastable atoms may not necessarily all be able to reach the detector. Only if the solid angle Ω_1 subtended by the detector, seen through the slit system from thecentre of the bombardment region, is larger than the solid angle Ω_2 of the beam spread due to recoil will the metastable atoms clear the intervening slit system. Otherwise only a fraction δ of these, called the 'recoil dilution' factor, may do so. One may then set

$$\delta = \begin{cases} \Omega_1/\Omega_2 & \Omega_1 < \Omega_2, \\ 1 & \Omega_1 \geqslant \Omega_2. \end{cases} \qquad \ldots\ldots(29)$$

The number of electrons per second ejected from the target due to metastable atoms is then

$$S = \frac{n_0 a v X A f \delta \mu \eta}{4\pi R^2}, \qquad \ldots\ldots(30)$$

where the factor μ expresses the fraction of metastables surviving their passage through any quenching electric fields, and η is the efficiency of the electron ejection by a metastable hydrogen atom from the metal surface used in the target.

The bombardment efficiency f may be estimated as follows: let the electron stream have I electrons per second, a height h and a width w. It takes a time w/v for an atom of speed v to pass through the bombardment region. The transition rate for an excitation process having a cross section σ is $I\sigma/wh$ so that the probability of excitation is

$$f=\frac{I\sigma}{hv}=2{\cdot}3\times10^{-8} \qquad \ldots\ldots(31)$$

(much less than unity!) with the typical values of bombarding current of 200 μa. $\sigma=10^{-17}$ cm^2, $h=1$ cm., $v=8\times10^5$ cm/sec., so that about one in forty million atoms is excited to the 2s state on passage through the bombardment region.

The detector signal is finally

$$S=n_0aXAI\sigma\delta\mu\eta/(4\pi R^2h) \text{ electrons per second.} \qquad \ldots\ldots(32)$$

Unfortunately many of the factors needed for the estimate of the yield were not known and a certain degree of optimism was required in the assumption of relatively favourable values for them in order to predict a usable signal. Typical values were:

$$I=1{\cdot}87\times10^{15} \text{ electrons/sec.}(0{\cdot}3 \text{ ma.}),\quad a=3{\cdot}1\times10^{-9} \text{ cm}^2,\quad A=1{\cdot}21 \text{ cm}^2,$$
$$n_0=2{\cdot}94\times10^{15} \text{ cm}^{-3}\quad(10^{-3} \text{ atm., } 2{,}500^\circ \text{K.}),\quad R=6{\cdot}35 \text{ cm.},\quad \sigma=10^{-17} \text{ cm}^2,$$
$$h=1 \text{ cm.},\quad X=0{\cdot}64,\quad \delta=0{\cdot}5,\quad \mu=0{\cdot}5,\quad \eta=0{\cdot}5 \qquad \ldots\ldots(33)$$

The values for a, A, R, h, I, n_0, X were at least near the actual values used. As explained above, the value for σ was based on very inadequate theory. The value $\mu=\frac{1}{2}$ for the metastable survival factor was reasonable (see § 5.13), provided quenching effects were not present, including those which might be due to contact potential differences, or to electric fields caused by charges trapped on insulating surface contaminations. The detector efficiency $\eta=0{\cdot}50$ was taken most optimistically. The value of the recoil dilution factor $\delta=\frac{1}{2}$ was obtained by considering the width of the distribution of recoil angles and the areas of the holes through which the deflected beam must pass. One may not take these openings too large, for then the size of the region over which uniform magnetic and radio-frequency fields must be maintained becomes excessive. Also, the background signal due to photons originating in the bombardment region is increased relative to the signal, since the photons are not restricted to a relatively small range of solid angles.

With the above assumptions the signal is

$$S=3{\cdot}26\times10^7 \text{ electrons/sec.}=5{\cdot}2\times10^{-12} \text{ amperes.} \qquad \ldots\ldots(34)$$

This is a current which is about 5×10^4 times the least current (10^{-16} ampere) which can be detected conveniently with an FP54 electrometer circuit and a sensitive galvanometer. Consequently, unless one or more of the estimates proved too optimistic, the signal should be large enough to detect, and even to use for accurate radio-frequency spectroscopy. As will be seen later, the overall estimate did prove to be very optimistic.

5.10. *Radio-Frequency Power Required*

The metastable atoms are to be subjected to radio-frequency waves somewhere between the source and detector. When the frequency is such that $\hbar\omega$ is equal or nearly equal to an energy difference between a Zeeman component of the $2^2S_{1/2}$ level and a Zeeman component of one of the 2^2P levels, the radiation may induce a transition to a non-metastable level and the detected signal will decrease. The amount of decrease will depend on the intensity of the radiation as well as its frequency, on the speed of the atoms, and the length of the radio-frequency field region. An estimate of the amount of radio-frequency power required for an appreciable quenching of the beam will now be given.

According to the quantum theory of radiation the decay rate of state $2^2S_{1/2}$ to one of the 2p states due to radio-frequency radiation is

$$\frac{1}{\tau_s} = \frac{2\pi e^2 \gamma S_0}{c\hbar^2} \frac{|(|\hat{\mathbf{e}}.\mathbf{r}|)|^2}{(\omega-\omega_0)^2+\frac{1}{4}\gamma^2}, \qquad \ldots\ldots(35)$$

where S_0 is the incident energy flux density of the radiation having circular frequency ω and electric polarization parallel to the unit vector $\hat{\mathbf{e}}$, ω_0 is the

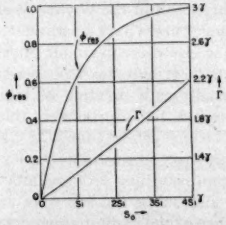


Figure 13. Effects of radio-frequency saturation. Resonance quenching and effective half-width as functions of radio-frequency power as given by equations (38) and (35). Hyperfine structure is not considered.

resonance circular frequency and $(|\mathbf{r}|)$ is the matrix element of the coordinate vector $\mathbf{r}$ for the transition in question. As before, $\gamma = 1/\tau_p$ is the radiative damping constant for the 2p states. At resonance $\omega = \omega_0$ and

$$\frac{1}{\tau_s} = \frac{8\pi e^2 S_0 |(|\hat{\mathbf{e}}.\mathbf{r}|)|^2}{c\hbar^2\gamma}. \qquad \ldots\ldots(36)$$

The matrix elements of $\hat{\mathbf{e}}.\mathbf{r}$ may be calculated from equations given by Bethe (1933, pp. 447–452), and for a transition from $2^2S_{1/2}$ $(m=\frac{1}{2})$ to $2^2P_{1/2}$ $(m=-\frac{1}{2})$ with $\hat{\mathbf{e}}$ along x

$$|(|x|)|^2 = 3a_0^2. \qquad \ldots\ldots(37)$$

The fraction of atoms quenched by the radio-frequency field is

$$\phi = 1-\exp\{-(l/v\tau_s)\}, \qquad \ldots\ldots(38)$$

where l is the length of the radio-frequency region and v is the speed of the atoms. For $l=1$ cm. and $v=8\times10^5$ cm/sec. the beam would be 63% quenched for

$$\tau_s = l/v = 1{\cdot}25\times10^{-6} \text{ sec.} \qquad \ldots\ldots(39)$$

According to equation (36) this requires an energy flux density at resonance of $S_0 = 3{\cdot}4$ mw/cm^2. Since this power density is easily obtained at any frequency

up to 30,000 Mc/s., no difficulty with radio-frequency power requirements was anticipated.

It is clear from equations (38) and (36) that when the radio-frequency power is sufficient to give nearly complete quenching at resonance the effective resonance curve will be appreciably broadened by radio-frequency saturation. The dependence of quenching and effective half-width Γ on radio-frequency power is shown in Figure 13.

5.11. *Breadth of Resonance Curves*

According to equation (35) the decrease in the beam will take place when $\omega = \omega_0$ and also for a band of frequencies around $\omega_0/2\pi$ of half-width $\gamma/2\pi$ or 99·8 Mc/s. This is just the uncertainty principle width ($\Delta E . \Delta t \sim \hbar$) of the decaying 2^2P states and cannot be reduced. In addition, as will be seen later, hyperfine structure increased the width of the lines by as much as a factor of two in some cases. The large width of the resonance curves is a significant fraction of the frequencies of the transitions and represents the greatest difficulty in the way of a really precise test of the theory. On the other hand, an opportunity is afforded for a quite accurate test of the Wigner–Weisskopf theory of radiation line shape, since the resonance curves have a natural width relative to the working frequency much greater than elsewhere in atomic physics.

The most obvious way to determine the resonance frequencies $\omega_0/2\pi$ would be to measure the beam intensity as a function of radio frequency, keeping the radio-frequency power constant. Unfortunately this is next to impossible to do. Microwave oscillators which can be tuned through hundreds of megacycles per second are available, but their output is frequency-dependent. Even if this were not the case the transmission line or waveguide to the quenching region would be electrically long and unless the line were very well broadbanded, the standing-wave ratio, and hence the radio-frequency voltage seen by the atom would vary with frequency. One might hope to monitor the radio-frequency power level with a crystal probe or bolometer mounted in the quenching region, but such devices are themselves frequency-sensitive.

Since a magnetic field was believed necessary anyway, Lamb and Retherford decided to take advantage of it to overcome the above difficulties. The frequency (and power level) of the oscillator is kept constant and the atomic energy levels are moved through resonance by varying the magnetic field. To a first approximation, the atomic energies and therefore frequencies are linear functions of magnetic field, so that a resonance curve taken in this way closely resembles the usual resonance curve. In order to interpret the result, however, it is necessary to know the values of the magnetic field and to have the theory of the Zeeman effect of the hydrogen fine structure levels for $n = 2$. The results of this theory are summarized in the next section.

5.12. *Zeeman Effect of the Fine Structure of Hydrogen*

For the present it will suffice to consider an approximate theory of the Zeeman effect of the hydrogen fine structure in which the interaction of the atom with an external magnetic field $\mathbf{H}$ is represented by the perturbing energy

$$\mathscr{H}' = \mu_0(\mathbf{L} . \mathbf{H} + 2\mathbf{S} . \mathbf{H}), \qquad \ldots\ldots(40)$$

where

$$\mu_0 = e\hbar/2mc \qquad \ldots\ldots(41)$$

is the Bohr magneton, $L=(\mathbf{r}\times\mathbf{p})/\hbar$ the orbital angular momentum and $\mathbf{S}$ the spin angular momentum of the electron measured in units of $\hbar$. The splitting of the $2^2S_{1/2}$ levels may be considered independently from that of the 2^2P levels for $\mathscr{H}'$ has no matrix elements connecting s and p states. The solution of this problem is well known and the results may be taken from Bethe's *Handbuch* article (1933, p. 396).

For $2^2S_{1/2}$, $m_s=\pm 1/2$

$$E(2^2S_{1/2}, m_s;\ H)=E(2^2S_{1/2})+2\mu_0 H m_s. \qquad \ldots\ldots(42)$$

For $2^2P_{3/2}$, $m_j=\pm 3/2$

$$E(2^2P_{3/2}, m_j=\pm\tfrac{3}{2};\ H)=E(2^2P_{3/2})+\tfrac{4}{3}\mu_0 H m_j. \qquad \ldots\ldots(43)$$

For the $m_j=\pm\frac{1}{2}$ levels of $2^2P_{1/2,\ 3/2}$ the levels are given by the roots of a secular determinant as

$$E=\tfrac{1}{2}(E_++E_-)+\mu_0 H m_j\pm\tfrac{1}{2}[(E_+-E_-)^2+\tfrac{4}{3}\mu_0 H(E_+-E_-)m_j+(\mu_0 H)^2]^{1/2}\ldots\ldots(44)$$

where E_+ and E_- are the zero field energies of $2^2P_{3/2}$ and $2^2P_{1/2}$ respectively.

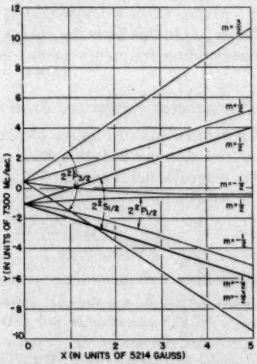


Figure 14. Zeeman splitting of the fine structure of states $n=2$ of hydrogen according to the Dirac theory.

If energy is measured in units of

$$h\nu_1=\tfrac{3}{5}(E_+-E_-)\simeq 7{,}300\text{ Mc/s.} \qquad \ldots\ldots(45)$$

from the centre of gravity $(2E_++E_-)/3$ and magnetic field in units

$$H_1=\frac{2(E_+-E_-)}{3\mu_0}\simeq 5{,}214\text{ gauss} \qquad \ldots\ldots(46)$$

by setting $\mu_0 H=\frac{3}{5}(E_+-E_-)x,$ $\ldots\ldots(47)$

$$E=\tfrac{3}{5}(E_+-E_-)y+\tfrac{1}{3}(2E_++E_-), \qquad \ldots\ldots(48)$$

one obtains $y=-\frac{1}{4}+m_j x\pm\frac{1}{2}(x^2+2m_j x+\frac{9}{4})^{1/2}.$ $\ldots\ldots(49)$

In the same notation the energy levels for $2^2P_{1/2}$ $m_j=\pm\frac{3}{2}$ are given by

$$y=\tfrac{1}{2}\pm 2x \qquad \ldots\ldots(50)$$

and for $2^2S_{1/2}$, $m_s=\pm\frac{1}{2}$, $y=y_0-1\pm x,$ $\ldots\ldots(51)$

where y_0 measures any displacement upward of $2^2S_{1/2}$ compared with $2^2P_{1/2}$ in zero magnetic field. A plot of these energy levels is given in Figure 14 showing the anomalous Zeeman, intermediate field, and Paschen–Back regions. Here $y_0=0$ as predicted by the Dirac theory. For the sake of brevity these energy levels $2^2S_{1/2}$ $m_s=\frac{1}{2}$ and $m_s=-\frac{1}{2}$ will be denoted by α and β respectively (as suggested by the conventional symbols for Pauli spin wave functions). The 2^2P levels will be devoted by a, b, c, d, e, f as shown in Figure 15 which enlarges the anomalous Zeeman and intermediate field regions of primary interest to us.

The selection rules for electric dipole radiation in the anomalous Zeeman region are $\Delta m_j=0$ for the electric vector of the radio waves parallel to the magnetic field, and $\Delta m_j=\pm 1$ for perpendicular polarization. Consequently, for parallel polarization the allowed transitions are αb, αe, βc, βf, while for perpendicular polarization the allowed transitions are αa, αc, αf, βb, βd, βe. The frequencies to be expected for these ten transitions are plotted in Figure 16 as a function of the magnetic field.

Actually, taking the fine structure separation $\frac{1}{16}\alpha^2cR$ as 10,950 Mc/s. and the s level shift $\delta E/h$ as 1,051 Mc/s., we now know that $y_0\sim 0{\cdot}144$ and not zero as predicted by the Dirac theory. On this basis the energy level curves are as shown in Figure 17. The corresponding frequency versus field curves are shown in Figure 18.

5.13. *Production of a Polarized Beam of Atoms*

It will be noted from a comparison of Figures 15 and 17 that the existence of the $2^2S_{1/2}$ level shift gives rise to a marked difference in the stability of the two metastable states α and β in magnetic fields around 575 gauss. For the quenching effect of an electric field $\mathbf{E}$ is given by

$$\frac{1}{\tau_s}=\gamma\sum_n \frac{|V_{ni}|^2}{\hbar^2[\omega_{ni}{}^2+\frac{1}{4}\gamma^2]}, \qquad \ldots\ldots(52)$$

where $V_{ni}=(n\,|e\mathbf{E}\cdot\mathbf{r}|\,i)$ is the matrix element of the perturbation $e\mathbf{E}\cdot\mathbf{r}$ connecting the two interacting states, $i=\alpha$ or β, $n=$ any 2P state and $\hbar\omega_{ni}$ is the energy difference between n and i. At 575 gauss $\omega/2\pi$ is 2,150 Mc/s. for the separation between states αf and zero for the states βe. As a result, the contribution to the quenching of the lower metastable state β by these states will be more rapid than for the upper state by a factor of

$$1+\frac{\hbar\omega_{\alpha f}{}^2}{\gamma^2}\simeq 1{,}850. \qquad \ldots\ldots(53)$$

The main known perturbing electric field is that due to the motion of the atoms at right angles to the magnetic field $\mathbf{H}$, and is given by

$$\mathbf{E}=(\mathbf{v}\times\mathbf{H})/c. \qquad \ldots\ldots(54)$$

Hence $\mathbf{E}$ is perpendicular to $\mathbf{H}$. If a right-handed Cartesian coordinate system with z along $\mathbf{H}$, x along $\mathbf{v}$ is introduced, then $\mathbf{E}$ is along y. The perturbation energy $e\mathbf{E}\cdot\mathbf{r}$ has matrix elements connecting states β and e, but not states β and f. For $v=8\times10^5$ cm/sec. and $H=575$ gauss, $E=4{\cdot}5$ volts/cm.

The life of the degenerate state β in a perturbing electric field of 4·5 volts/cm. perpendicular to a magnetic field of 575 gauss is given by equation (52) to be $3{\cdot}8\times10^{-8}$ sec. while state α has a life of 7×10^{-5} sec. This means that in the

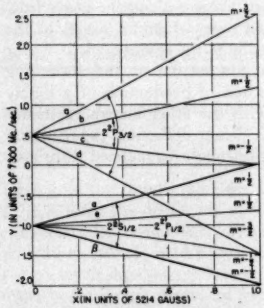


Figure 15. Enlargement of Figure 14 showing the abbreviations used for the Zeeman component energy levels.

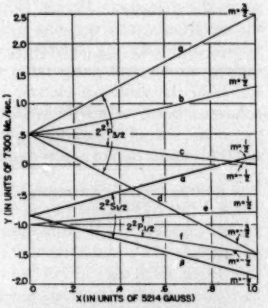


Figure 17. Zeeman energy levels as in Figure 14, but with the $2^2S_{1/2}$ pattern raised by about 1,000 Mc/s.

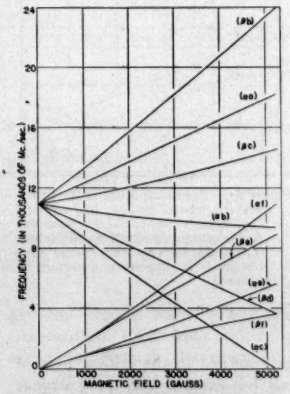


Figure 16. Resonance frequencies to be expected as functions of magnetic field for all allowed transitions leading from the metastable states α ($2^2S_{1/2}$, $m=\frac{1}{2}$) and β ($2^2S_{1/2}$, $m=-\frac{1}{2}$) to the non-metastable states a, b, c, d ($2^2P_{3/2}$) and e, f ($2^2P_{1/2}$) according to the Dirac theory.

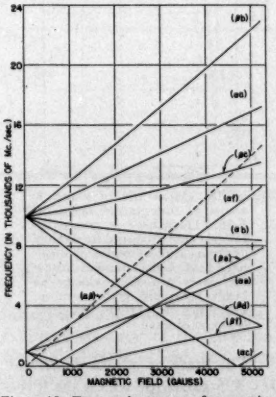


Figure 18. Expected resonance frequencies as functions of magnetic field as in Figure 16, but with the $2^2S_{1/2}$ states raised by about 1,000 Mc/s. The frequency of the transition $\alpha\beta$ between the metastable states is also shown.

presence of this motional electric field, atoms in state α will nearly all be able to keep their excitation while they travel 6 cm. to a detector, while practically none of the atoms originally in the lower metastable state β will be excited when they reach the detector. Hence, after moving a short distance along the beam from the electron bombarder, one will have a polarized beam of atoms, nearly all the excited atoms having electron spins oriented parallel to the magnetic field. The excited atoms with antiparallel spins have been filtered out by the quenching action of the motional electric field. This method of production of a highly polarized beam of matter is to be contrasted with the other known methods: the small spatial separations obtained in atomic beam deflection experiments, and the polarization of neutrons on passage through strongly magnetized absorbers.

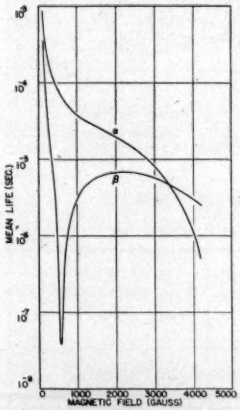


Figure 19. Mean life of α and β states as function of magnetic field. Quenching is caused by electric field produced by motion of atoms at right angles to magnetic field. Calculations are made for most probable velocity in beam emerging from oven at 2,500° K. Hyperfine structure is not considered.

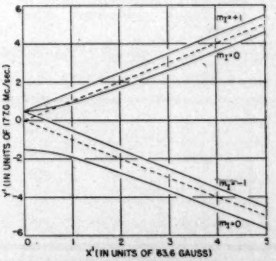


Figure 20. Zeeman splitting of the hyperfine structure of $2^2S_{1/2}$. The zero field separation is taken to be $\frac{1}{4}$ that measured for the ground state by Nafe and Nelson. The dotted lines show the energy levels obtained if hyperfine structure is ignored.

The state of affairs described above is not confined to the critical field of $H=575$ gauss for which the levels β and e cross, but extends over quite a range of field strengths. A plot of life against magnetic field is given in Figure 19. With the apparatus to be described in this paper, transitions βb, βc, βd, βe, βf of Figure 18 could be observed, but had very small intensities.

Motional Stark quenching also occurs about the magnetic field where levels α and c cross or $H=4{,}650$ gauss. This region is much wider than for the βf crossing because the motional electric field is larger and the beam of metastable atoms will be strongly quenched above 3,000 gauss.

5.14. *Effect of Hyperfine Structure on Energy Levels*

The discussion of the energy levels of the $n=2$ states in §5.12 ignored the interaction of the magnetic moment of the nucleus (proton, deuteron, ...) with the electron. Fortunately, although the effects of this are not negligible by any means, they can be taken into account as an afterthought by perturbation theory for the purposes of the present discussion. The magnetic moment of the nucleus

$$\boldsymbol{\mu} = g_I \mu_0 \mathbf{I} \qquad \ldots\ldots(55)$$

produces a magnetic field derivable from the vector potential

$$\mathbf{A} = \frac{\boldsymbol{\mu} \times \mathbf{r}}{r^3}. \qquad \ldots\ldots(56)$$

Here g_I is the Landé value for the nucleus (about 5·6/1,836 for the proton and 1·7/1,836 for the deuteron), μ_0 is the Bohr magneton (equation (41)) and $\mathbf{I}$ is the spin vector for the nucleus. According to Dirac's equation, the interaction energy with the electron is $e\boldsymbol{\alpha}.\mathbf{A}$ where $\alpha_x, \alpha_y, \alpha_z$ are the Dirac matrices. The shift of the energy levels due to hyperfine structure is small and may be calculated using first order perturbation theory by forming the average of $e\boldsymbol{\alpha}.\mathbf{A}$ over the unperturbed Dirac four component wave function for the state of the electron. In forming this average the two small component wave functions must be calculated with some care near the nucleus in order to avoid ambiguities due to the singularity at the origin in $\mathbf{A}$. The theory is conveniently worked out by Bethe (1933, p. 385) for the case of vanishing external magnetic field. The results for the energies are given below.

For s states the effective perturbation operator is

$$w = \frac{16\pi}{3} g_I \mu_0^2 |\psi(0)|^2 \mathbf{I}.\mathbf{S} \qquad \ldots\ldots(57)$$

or

$$w = \frac{8}{3} g_I \frac{Z^3}{n^3} \alpha^2 hcR \mathbf{I}.\mathbf{S} \qquad \ldots\ldots(58)$$

giving energy shifts

$$w = \frac{4}{3} g_I \frac{Z^3}{n^3} \alpha^2 hcR. \begin{cases} I & \text{for } F = I + \tfrac{1}{2}, \\ -(I+1) & \text{for } F = I - \tfrac{1}{2}. \end{cases} \qquad \ldots\ldots(59)$$

The separation of the two states is

$$\Delta w_n = \frac{8}{3} g_I (I + \tfrac{1}{2}) \frac{Z^3}{n^3} \alpha^2 hcR. \qquad \ldots\ldots(60)$$

For the ground state of hydrogen, this splitting amounts to 1,416 Mc/s., (increased to 1,420 Mc/s. by the anomalous magnetic moment of the electron as shown through the work of Nafe and Nelson (1948)). For the case $n=2$, of interest here, the splitting is one-eighth as much, or 177 Mc/s.

The theory of the Zeeman effect of this hyperfine structure for all magnetic fields is given by the well-known Breit–Rabi (Breit and Rabi 1931) formula and is illustrated in Figure 20. For the present purposes it suffices to consider only the strong field limit in which $\mathbf{I}$ and $\mathbf{S}$ are decoupled from one another and precess independently around $\mathbf{H}$ with projections m_I and m_s respectively. Then $\mathbf{I}.\mathbf{S}$ can be replaced by $m_I m_s$ and one obtains for the hyperfine energy contribution

$$w = \frac{8}{3} g_I \frac{Z^3}{n^3} \alpha^2 hcR m_I m_s \qquad \ldots\ldots(61)$$

State α, for which $m_s = +\frac{1}{2}$ is therefore split into $(2I+1)$ states. For hydrogen, with $I=\frac{1}{2}$, there are two states separated by $\frac{1}{2}\Delta w$ or 88 Mc/s. State $\beta(m_s=-\frac{1}{2})$ is similarly split, but with an inversion because of the sign change of m_s. In the interpretation of the more precise measurements, it is, of course, necessary to question the validity of the strong field approximation. For the present purposes the error is not large since the dimensionless parameter

$$x' = (g_J - g_I)\mu_0 H/\Delta w \qquad \ldots\ldots(62)$$

reaches unity at a magnetic field of only 63·6 gauss, and nearly all of the data described here were taken at fields well above this.

For other than s states Bethe assumed that the fine structure splitting was large compared with the hyperfine structure separation and derived the equation

$$w = 2\mu_0{}^2 g_I \left(\frac{1}{r^3}\right)_{\text{av}} \frac{L(L+1)\mathbf{I}.\mathbf{J}}{J(J+1)} = g_I \frac{Z^3}{n^3}\frac{\alpha^2 hcR\mathbf{I}.\mathbf{J}}{(L+\frac{1}{2})J(J+1)} \qquad \ldots\ldots(63)$$

for the perturbation operator. In zero magnetic field

$$\left.\begin{aligned} \mathbf{I}.\mathbf{J} &= \tfrac{1}{2}[F(F+1) - I(I+1) - J(J+1)] \\ \text{and} \\ (\mathbf{I}.\mathbf{J})_{F=|I+J|} - (\mathbf{I}.\mathbf{J})_{F=|I-J|} &= \begin{cases} I(2J+1) & J \geqslant I \\ J(2I+1) & J < I \end{cases} \end{aligned}\right\} \qquad \ldots\ldots(64)$$

so that for hydrogen the state $P_{1/2}$ splits into two states with resultant angular momentum quantum number $F=1, 0$ and $P_{3/2}$ splits into two states with $F=2, 1$. The separations are respectively one-third and two-fifteenths as much as for the corresponding $2S_{1/2}$ state.

In the presence of a magnetic field which is not large enough to break down the coupling between **L** and **S** to form a resultant **J**, one may continue to use the above equation (63). However, one must add to the Hamiltonian the energy of orientation of the vectors **I** and **J** in the magnetic field

$$\mu_0(g_J\mathbf{J} - g_I\mathbf{I}).\mathbf{H}$$

and the problem becomes mathematically similar to that of the Zeeman effect of the hyperfine structure for s states, so that the Breit–Rabi formula may be used if either I or J is $\frac{1}{2}$.

A simplification may be obtained if the magnetic field is large enough to decouple **I** from the vector **J**, but not sufficient to decouple **L** and **S**. The $\mathbf{I}.\mathbf{J}$ may be replaced by $m_I m_J$ and the hyperfine energy contribution becomes

$$w = \frac{g_I Z^3 \alpha^2 hcR m_I m_J}{n^3 J(J+1)(L+\frac{1}{2})} + \mu_0 H(g_J m_J - g_I m_I). \qquad \ldots\ldots(65)$$

As a result of the hyperfine structure, again taking the case $I=\frac{1}{2}$ for simplicity, each state is in effect split into two states equally removed in energy from the unperturbed position. To the approximation considered here, the hyperfine splittings are dependent on magnetic field only because of the last term—$\mu_0 H g_I m_I$ of equation (65). The splittings are expected to be $\frac{1}{2}\Delta w = 88$ Mc/s. for the $2^2S_{1/2}$ states, $\frac{1}{6}\Delta w = 29$ Mc/s. for the $2^2P_{1/2}$ states, while it is $\frac{1}{30}\Delta w = 6$ Mc/s. for the $m_J = \pm\frac{1}{2}$ states of $2^2P_{3/2}$ and $\frac{1}{10}\Delta w = 18$ Mc/s. for the $m_J = \pm 3/2$ states of $2^2P_{3/2}$. For the case of deuterium, where $I=1$ the patterns are more complicated, but the splittings are very much smaller because of the smaller nuclear magnetic moment.

The splitting of the energy levels due to hyperfine structure gives rise to a complication of the observed resonance curves. In most cases the natural width of the curves exceeds the splitting, and the result is a composite of two more or less imperfectly resolved resonance curves of the Wigner–Weisskopf type. The case of transitions from state α ($2^2S_{1/2}$, $m_s=\frac{1}{2}$) to states e ($2^2P_{1/2}$, $m_J=\frac{1}{2}$) and f ($2^2P_{1/2}$, $m_J=-\frac{1}{2}$) will serve to illustrate this. The observed transitions are electric dipole in type, and as the nuclear moment is decoupled from the other

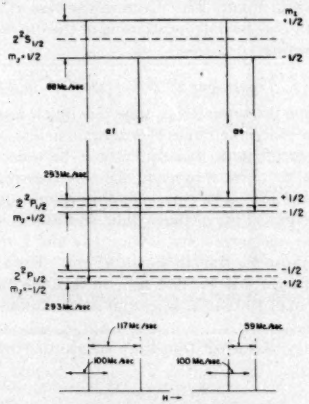


Figure 21. Hyperfine structure of the $2^2S_{1/2}$ ($m=\frac{1}{2}$) and $2^2P_{1/2}$ levels in strong magnetic field. The allowed transitions are indicated by the arrows. The resonance curve peaks αf and αe are separated by 117 and 59 Mc/s. respectively. This separation is to be compared with the radiative width of 100 Mc/s. For simplicity, the hyperfine components have been taken to be equally separated from the dotted position, corresponding to complete Back–Goudsmit decoupling of I and J.

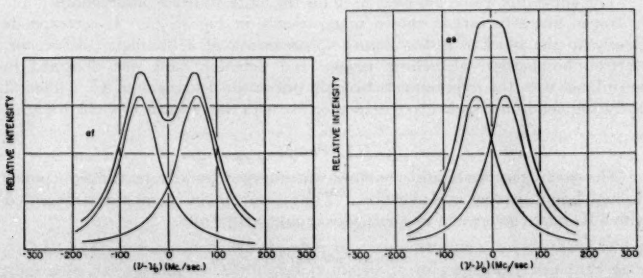


Figure 22. Ideal theoretical resonance curves showing the effect of hyperfine structure for transitions αf and αe respectively. When plotted as functions of magnetic field, the resonance curve for αf becomes the narrower of the two.

angular momentum vectors by the magnetic field, the selection rule $\Delta m_I = 0$ obtains.

The allowed transitions are shown in Figure 21. It will be noted that because of the reversal of sign of the product $m_I m_J$ the separation of the two peaks is $4\Delta w/3$ or 117 Mc/s. for transition αf and only $2\Delta w/3$ or 58 Mc/s. for transition αe. The natural width of the separate transitions is 100 Mc/s. When two such curves are superposed the resulting complex has two peaks only if the component peaks are separated by half the natural width or by 50 Mc/s. The theoretical forms of such curves are shown in Figure 22. Even in the case αf, where the separation is 177 Mc/s., the result may be only a flattening of the top of the curve if there are other causes of broadening.

5.15. *Transitions* $\alpha\beta$ $2^2S_{1/2}$ $(m_s = \frac{1}{2} \rightarrow m_s = -\frac{1}{2})$

As mentioned above the lower $2^2S_{1/2}$ state β is much less metastable than the upper state α for a considerable range of magnetic fields. Atoms in the β state do not reach the detector and consequently it becomes possible to observe transitions from α to β. The frequency for this transition is plotted against magnetic field in Figure 18. The corresponding resonance curves have a width determined by the reciprocal life of the β state, and as this is considerably longer than the life of the non-metastable states such as e and f, the αβ resonances are much narrower than those for transitions αe, αf etc. Even at the crossing point of β and e, the life of state β is 4×10^{-8} sec., so that the half-width ought to be $1/2\pi\tau$ or 4 Mc/s. instead of 100 Mc/s. as for transitions to the more highly damped states.

Since the frequency of the αβ transition (neglecting hyperfine structure) is given by

$$\hbar\omega_{\alpha\beta} = g_s\mu_0 H, \qquad \ldots\ldots (66)$$

a very convenient method for determination of magnetic field is available. A small correction for Stark effect associated with the motional electric field must be made. Some examples of the sharp αβ resonances are given later in this report.

§6. APPARATUS

The apparatus which has been used for the more accurate observations of the hydrogen fine structure is shown schematically in Figure 23. It corresponds closely to the block diagram Figure 4, consisting of a hydrogen dissociator, electron bombarder, quenching region and detector, and was designed in accordance with the requirements brought out in the discussion of §5. Typical operating conditions were given in connection with the estimate of yield in §5.9.

6.1. *Pumps*

The hydrogen dissociator chamber was pumped by a large diffusion pump through liquid nitrogen cooled baffles. The pressure in the chamber was measured with a Knudsen gauge and was normally about 6×10^{-4} mm.

6.2. *Gas Supply*

The hydrogen or deuterium was taken from a commercially available tank provided with a gas regulator and passed through hot palladium and a variable leak made by deforming a flattened copper tube.

6.3. *Hydrogen Dissociator*

The tungsten oven and water-cooled leads for the heating current are indicated in Figure 23. As estimated in § 5.5 the hydrogen pressure inside the oven was about 0·7 mm. It required 300 watts of heater power to reach the operating temperature of 2,500–2,600° K. The increased life of the tungsten oven at these temperatures compensated for the relatively inefficient atomic yield. The horizontal and vertical position of the hydrogen dissociator could be adjusted to maximize the signal.

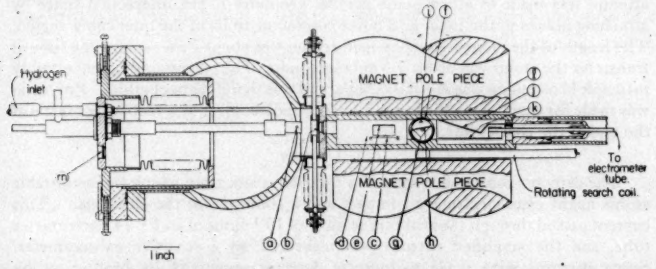


Figure 23. Cross section of apparatus used by Lamb and Retherford for determination of hydrogen fine structure.

a. Tungsten tube for hydrogen dissociation.
b. Adjustable slit system.
c. Electron bombarder cathode.
d. Electron bombarder grid.
e. Electron bombarder plate.
f. Quenching region.
g. Openings for passage of metastable atoms through quenching region.
h. Plate attached to centre conductor of radio-frequency transmission line.
i. Electrode for application of static electric field to beam of metastable atoms.
j. Target for metastable atoms.
k. Collector for electrons ejected from target by metastable atoms.
l. Pole face of magnet shaped to increase homogeneity of magnetic field in quenching region.

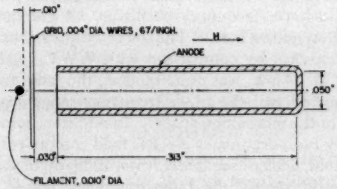


Figure 24. Cross section of electron gun showing parallel metal plates for reduction of space charge effects.

6.4. *Electron Bombarder*

The electron bombardment chamber was joined to the oven chamber by an adjustable slit opening. Differential pumping of this chamber was provided. Details of the electron bombarder are shown in Figure 24. The horizontal plates attached to the anode were found to be very helpful for the reduction of space charge effects. Since charged films of pump oil on the anode were particularly

troublesome, provision was made for heating this electrode. Even so, it was necessary to resort to mechanical cleaning of the anode at intervals of two or three weeks in order to maintain a good signal.

6.5. *Quenching Region*

The atoms excited to the $2^2S_{1/2}$ state could be subjected either to static electric or radio-frequency fields in the quenching region. This consisted of a vertical coaxial transmission line intersecting the line of flight of the excited atoms. An attempt was made to attain plane parallel geometry in the interaction space by attaching planes to the inner and outer conductor to form the quenching region. The length of the radio-frequency field region was about 1 cm., so that the time of transit for the atoms was about 1·6 microsecond. The corresponding uncertainty principle broadening was about 0·6 Mc/s. and was therefore negligible. Provision was made for monitoring the radio-frequency power level by a rectifying crystal at the end of the coaxial line.

6.6. *Detector*

The detector consisted simply of a tungsten plate from which the metastable atoms might eject electrons on impact and a collector for the electrons. This current passed through the grid bias resistor of 10^{11} ohms of an FP 54 electrometer tube, and the amplified current was measured on a sensitive galvanometer. Some attempts were made to increase detector sensitivity by heating, or by evaporating films of low work function metals on to it, but these did not meet with success. The detector chamber was also differentially pumped and the pressure during operation was of the order of 3×10^{-5} mm.

6.7. *Radio-frequency Techniques*

The frequencies used in observations of the hydrogen fine structure ran from 500 Mc/s. to 11,500 Mc/s. The oscillators included triodes, klystrons and continuous wave magnetrons. The radio-frequency power was turned on and off with a coaxial switch, or in some cases with a waveguide switch. The radio frequency was measured to a few hundredths of a megacycle per second by mixing with a harmonic of a lower frequency oscillator. A communications receiver measured the frequency of the beat. The lower frequency oscillator was crystal controlled and was checked by comparison with WWV. Although the power density requirement of 3·4 mw. was quite modest, the radio-frequency line could not very well be matched over the whole frequency range if plane parallel fields were to be achieved in the interaction space. In addition, some of the transitions were induced only by radio-frequency electric field components at right angles to the static magnetic field, and consequently it was difficult, in some cases, to obtain oscillators which could supply the radio-frequency power required and yet operate in a stable manner with little attenuation between oscillator and switch.

6.8. *Magnetic Field*

The apparatus was placed between the poles of a large electromagnet whose pole faces were shaped to increase the homogeneity of the field in the quenching region. The current producing the magnetic field was stabilized electronically to better than one part in 10^4. The magnetic field in the quenching region could be determined from a frequency measurement by observation of the m_s transitions $\alpha\beta$.

It was not convenient to observe these at more than a few frequencies. Consequently, a flip coil was rotated in the magnet gap by a synchronous motor, as shown in Figure 25. On the other end of the shaft a second flip coil was rotated in the field of a permanent magnet. The outputs of the two flip coils were balanced by use of a resistance bridge, the unbalance was amplified by a twin-T narrow-band amplifier tuned to the synchronous frequency and the output was observed on an oscilloscope. The permanent magnet was calibrated during each day's run by observation of the m_s transitions and the calibration was found to remain sufficiently constant from day to day. The magnetic field could then be measured in terms of the resistance required to balance the bridge. The field could be read to a few hundredths of a

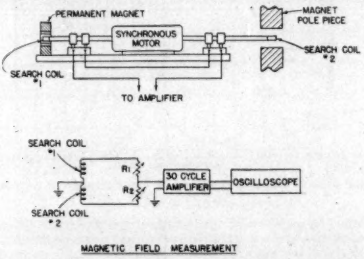


Figure 25. Apparatus for measuring magnetic fields.

gauss. Although the magnetic fields in the quenching region and at the first flip coil were not quite equal, the difference was found to be a linear function of the magnetic field, and hence only resulted in a slightly different value of the calibration constant in ohms per gauss for the field in the interaction space. Corrections were made for flip coil resistance changes due to ambient temperature variation. A slight correction for eddy current heating of the shaft was also necessary.

It was found that an A.C. magnetic field component of about a third of a gauss resulted from currents used to heat the tungsten oven and electron bombarder filament. This field could be cancelled out by passing current in a loop about one of the pole pieces.

§ 7. OBSERVATIONS ON HYDROGEN FINE STRUCTURE

7.1. *Procedure*

Under typical operating conditions chosen for a study of the resonances the galvanometer deflection due to the electron bombardment amounted to some 200 cm. (about 10^{-14} ampere). Of this, perhaps 130 cm. could be quenched by application of a large static electric field in the interaction space and corresponded to the signal of metastable hydrogen atoms. The background signal of 70 cm. was due to ultra-violet light produced in the bombardment region. If instead, radio-frequency power was introduced, a fraction of the signal was quenched, depending on the power and the amount by which the magnetic field departed from a resonance value. The results were expressed in terms of the percentage of the quenching of the metastable atom signal.

7.2. *Resonance Curves*

Examples of resonance curves taken at about 2,200 Mc/s. for transitions αe and αf are shown in Figures 26 and 27. The first of these Figures, obtained with hydrogen, shows the hyperfine structure resolved for the peak αf in accordance with the discussion of §5.14. The curves for deuterium in Figure 27 are consider-

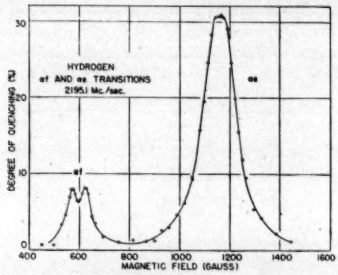


Figure 26. Panoramic view of resonances αf and αe for hydrogen at a frequency of 2,195·1 Mc/s.

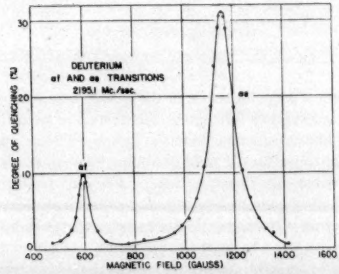


Figure 27. Panoramic view of resonances αf and αe for deuterium at a frequency of 2,195·1 Mc/s.

ably sharper, and no hyperfine structure is resolved. The radio-frequency power was adjusted so that the peak quenching was equal in both cases for transition αe. The relative weakness of peaks αf is due to the radio-frequency electric field configuration, which favours transitions with $\Delta m_j = 0$.

7.3. *Symmetry of Resonances*

Before data of this type can lead to a highly precise value of the shift of the $2^2S_{1/2}$ level, it is necessary to determine how nearly the peaks of the resonance curves, or alternately, the centres defined in some other manner, correspond to the actual positions of the energy levels. Any asymmetry or shift could lead to an error in the result. Among the factors which might contribute to this are (1) possible asymmetries or shifts basic to radiation theory, (2) overlap of nearby resonances, (3) variation of the matrix element for the transition as the magnetic

field is varied, (4) incomplete Paschen–Back effect for the hyperfine components, (5) differential quenching of the hyperfine components by motional electric fields, (6) curvature of the Zeeman energy curves, (7) Stark effects, (8) presence in the beam of a small fraction of atoms in the β state. A lengthy programme of calculations and measurements is required to allow for all such sources of error. For present purposes, however, it suffices to take a more empirical point of view. The

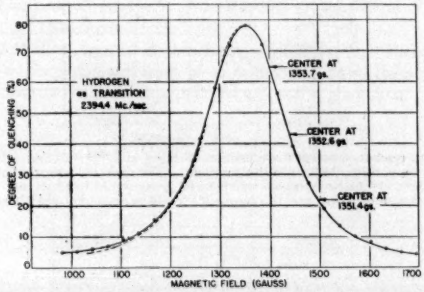


Figure 28. Resonance αe for hydrogen taken with a large excess of radio-frequency power at 2,394·4 Mc/s. The dotted curve at the left indicates the mirror image of the right-hand part of the curve about an arbitrary magnetic field. The centres determined at different levels differ by 1–2 gauss, and would lead to values of the level shift differing by 1–2 Mc/s.

resonance curves are actually highly symmetrical. An example of this is shown in Figure 28 in which an excessive amount of power is used in order further to broaden the resonance and thereby increase the chance to observe an asymmetry. The centres of the curve obtained at different levels of quenching agree quite closely, and also with the centre obtained with much less power.

A further empirical check may be made by comparing the centres for different resonances αe and αf at various frequencies. The conclusion is that no corrections are needed to obtain a result accurate to $\pm$ 5 Mc/s. from uncorrected data.

7.4. *Results*

The result obtained in this way was a shift upward of $2^2S_{1/2}$ relative to $2^2P_{1/2}$ of 1,062 $\pm$ 5 Mc/s. for both H and D. This figure appears to be in definite disagreement with the currently calculated value of 1,051 Mc/s. The location of the $2^2P_{3/2}$ level is found to agree with theory within the above precision, if the anomalous magnetic moment of the electron is considered.

7.5. *Sharp Resonances*

As explained in §5.15, the sharp $\alpha\beta$ transitions were used to calibrate the magnetic field. These have a markedly different character depending on whether the level β is well separated from competing levels or not. An example of the first type is given in Figure 29, and some δf of the second type are shown in Figures 30 and 31. The resonances of second type are superposed on the broad resonance αe, as shown in Figure 32.

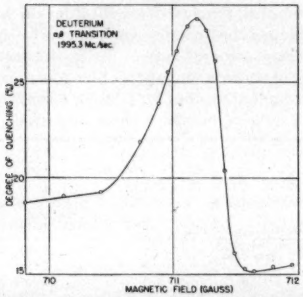


Figure 29. Sharp resonance curves for transition $\alpha\beta$ taken at 1,995·3 Mc/s. The peak shown is the centre one of three observed with deuterium. The peaks are separated by about 10 gauss. For hydrogen two peaks separated by about 63 gauss would be observed. The resonance fields for transitions αf and αe at this frequency would be about 490 and 970 gauss respectively.

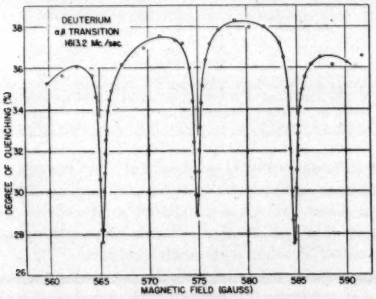


Figure 30. Sharp resonance curves for transitions $\alpha\beta$ in deuterium taken at the frequency 1,613·2 Mc/s. for which the broad resonance αe overlaps the sharp resonances.

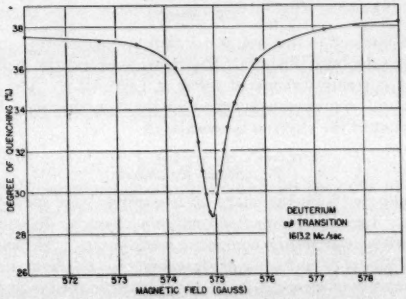


Figure 31. Detail of the central peak shown in Figure 30.

One of the sharp resonances is observed for each hyperfine state: two for hydrogen and three for deuterium. A peculiar feature of these transitions is that the quenching is *reduced* when the resonance condition $\omega = \omega_{\alpha\beta}$ is satisfied. In the first type of resonance this reduction occurs at the minimum on the right of the peak. The maximum is located essentially at the field to be expected if motional Stark effect shifting of the levels is considered. In the second type the quenching of each hyperfine component is also reduced to a small value when $\omega = \omega_{\alpha\beta}$. The width of the resonance corresponds to the lifetime of the β state at the crossing point of β and e as explained in § 5.15.

The theory of these transitions has been worked out (see Part III of paper cited in footnote on p. 20), and leads to an understanding of the above features. It is rather complicated, so that the equations will not be given here. However, it

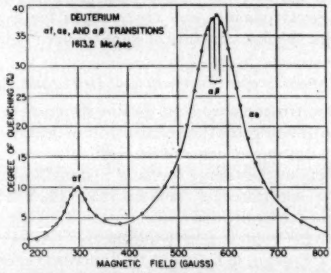


Figure 32. Resonance αe in deuterium taken at 1,613·2 Mc/s. to show setting for sharp resonances of Figure 30.

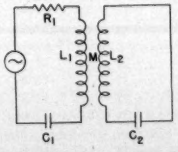


Figure 33. Electrical circuit analogue for interference of transitions αe and αβ.

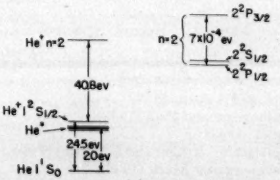


Figure 34. Energy levels of He and He+.

is possible to obtain a physical picture in terms of an electrical analogy with the problem of two coupled resonant circuits shown in Figure 33, consisting of a primary with parameters L_1, C_1, R_1 and a loss-less secondary with L_2, C_2 coupled to the first by a mutual inductance M. The resonance frequency $\omega_1 = 1/\sqrt{(L_1 C_1)}$ of the first circuit corresponds to the transition from α to e, and the radiation damping is represented by the resistive element R_1. The resonance $\omega_2 = 1/\sqrt{(L_2 C_2)}$ of the second circuit occurs at the frequency of the transition $\alpha\beta$. The β state does not decay directly, and this is represented by setting $R_2 = 0$. There is no direct coupling from α to β (except by relatively unimportant magnetic dipole radiation), while in the electrical circuit analogy, the second resonator is excited only indirectly by its

mutual inductive coupling to the first circuit. In the atomic case the motional Stark effect provides the coupling between β and e. The input admittance of such a circuit (and hence the power absorbed) has a broad resonance for $\omega = \omega_{\alpha e}$, while in a very narrow band about $\omega = \omega_{\alpha\beta}$ the reflected impedance of the loss-less secondary is large and reduces the response of the network.

§8. FINE STRUCTURE OF SINGLY IONIZED HELIUM

8.1. *Theoretical Expectations for He II Level Shifts*

According to equation (15) for the energy difference between the $2^2S_{1/2}$ and $2^2P_{1/2}$ levels, the theoretical shift for singly ionized helium can be found by putting $Z=2$ and $k_1(He^+) = 4k_1(H)$. Bethe (1947) predicted in this way that the level shift should be about 13 times that for hydrogen, or 13,820 Mc/s. In view of the discrepancy between calculations and observations for hydrogen it became of interest to re-examine the fine structure of singly ionized helium.

8.2. *Spectroscopic Observations of He II Fine Structure*

The history of spectroscopic work on the fine structure of He II is again a long one. Paschen (1916) investigated the fine structure of the line $\lambda 4684$ ($n=4 \rightarrow n=3$). His results constituted strong evidence of the essential correctness of the theoretical fine structure splitting of the energy levels of hydrogen-like atoms predicted by Sommerfeld. Later spectroscopic workers (Leo 1926, 1937, Paschen 1927, Chu 1939) were about evenly divided in their conclusions concerning the existence of a discrepancy from the Sommerfeld–Dirac theory. After the definite establishment of the $2^2S_{1/2} - 2^2P_{1/2}$ shift in hydrogen by a microwave method, a considerable amount of more accurate spectroscopic work was done in the visible and ultra-violet regions on the fine structure of He II. These results are summarized in Table 2.

Table 2. Recent Spectroscopic Observations on Fine Structure of Ionized Helium

Observers	δE_n (cm^{-1})	δE_2 (in 10^3 Mc/s.)
Mack and Austern (1947)	$0{\cdot}113 \pm 0{\cdot}014$ ($n=3$)	$11{\cdot}4 \pm 1{\cdot}4$
Kopfermann and Paul (1948)	$0{\cdot}137 \pm 0{\cdot}015$ ($n=3$)	$13{\cdot}9 \pm 1{\cdot}5$
Fowles (1948)	$0{\cdot}445 \pm 0{\cdot}050$ ($n=2$)	$13{\cdot}3 \pm 1{\cdot}5$
Kopfermann, Krüger and Öhlmann (1949)	$0{\cdot}119 \pm 0{\cdot}003$ ($n=3$)	$12{\cdot}0 \pm 0{\cdot}3$
Hirschberg and Mack (1950)	$0{\cdot}125 < \delta E_3 < 0{\cdot}134$ ($n=3$)	$12{\cdot}6 < \delta E_2 < 13{\cdot}6$
Murakawa, Suwa and Kamei (1949)*	$0{\cdot}134 \pm 0{\cdot}005$ ($n=3$)	$13{\cdot}6 \pm 0{\cdot}5$

* and unpublished work in 1950.

Fowles measured δE_2 by analysis of the ultra-violet line $\lambda 1640$ ($i=3 \rightarrow n=2$), while the others determined δE_3 by a study of the visible line $\lambda 4686$ ($n=4 \rightarrow n=3$). To facilitate comparison with the theoretical electromagnetic level shift, and later with the result of the microwave measurements, the value in the last column represents in each case the corresponding shift of the $n=2$ level, obtained by assuming a negligible change in $k_1(n, 0)$ when n changes from 3 to 2, in which case, according to equation (15), $\delta E_2/\delta E_3 = 27/8$. It is to be noted that the last two more accurate measurements are then somewhat low compared with the theoretical value of 13,820 Mc/s.

8.3. *Microwave Method for Study of He II Fine Structure*

The possibility of applying microwave techniques to a measurement of the fine structure of He^+ was investigated by Lamb and Skinner (1950). The theoretical value of 13,820 Mc/s. ($\lambda = 2{\cdot}2$ cm.) for the shift of the $2^2S_{1/2}$ state above that $2^2P_{1/2}$ lies in a readily accessible range of the microwave spectrum. The power required is $2^6 = 64$ times that needed to produce the same decay rate at resonance in hydrogen and was barely available. On the other hand, the doublet separation $2^2P_{3/2} - 2^2P_{1/2}$ corresponds to a wavelength of 1·7 mm. for which oscillators are not yet readily available.

As in the case of hydrogen, the first problem is to obtain a sufficient number of helium ions in an excited state. The difficulties with the use of a discharge tube method are about the same as before. A measurement by another method is possible only because of the metastability of the $2^2S_{1/2}$ state. The energy levels of interest are shown in Figure 34. An energy of 24·6 ev. is required to remove one electron from the helium atom forming the ground state $1^2S_{1/2}$ of the ion. An additional 40·8 ev. is required to excite the remaining electron to the $n = 2$ group of states. The lifetime of the 2^2P states is

$$\tau' = \frac{1}{\gamma'} = 10^{-10}\,\text{sec.} \qquad \ldots\ldots(67)$$

which is $2^4 = 16$ times smaller than the corresponding life for hydrogen. The $2^2S_{1/2}$ state is metastable. As with hydrogen, the most probable decay mechanism for an isolated ion is by double quantum emission. The decay rate for this process is $2^6 = 64$ times that calculated for hydrogen by Breit and Teller, so that the life of the $2^2S_{1/2}$ state should be $2{\cdot}2 \times 10^{-3}$ sec.

The energy difference between the $2^2S_{1/2}$ and $2^2P_{1/2}$ states can in principle be measured by obtaining a sufficient number of helium ions in the metastable $2^2S_{1/2}$ state, irradiating these with microwave radiation of the frequency required to induce transitions to the $2^2P_{1/2}$ state, and in some manner detecting the occurrence of the transition.

In contrast to the case of hydrogen it is difficult to form a beam of metastable helium ions. Even if one could do so the detection scheme used for hydrogen would not be satisfactory. It is known that ions may eject electrons from metal surfaces, so that there would be a large background detector signal due to the helium ions in the ground state which could not be separated from the much smaller number of metastable ions.

For this reason Lamb and Skinner adopted another method. Helium gas was bombarded by electrons of several hundred volts energy. The various excitation processes which occur in the bombardment region are accompanied by the emission of ultra-violet radiation and may be detected photoelectrically. This radiation consists of (*a*) photons of about 20–24·6 volts emitted in transitions from excited states to the ground state of He I, (*b*) radiation of energy 41–54·4 ev. emitted in transitions from excited states to the ground state of the ion He II, and (*c*) visible radiation emitted in other transitions. Of primary interest here is the radiation of 41 ev. energy emitted in transitions $n = 2 \rightarrow n = 1$ of the ion. This may arise either from states $2^2P_{1/2}$ or $2^2P_{3/2}$. The $2^2S_{1/2}$ state, being metastable, would not ordinarily contribute to the 41-volt radiation. If quenching fields are present or if transitions from 2s to 2p are induced in some way, an additional amount of the

41-volt radiation may be produced. The experiment consists of a detection of the increase in this radiation when such transitions are induced by radio waves of a frequency corresponding to the energy difference $2^2S_{1/2} \rightarrow 2^2P_{1/2}$.

Since the efficiency of photoelectric emission is quite high in the ultra-violet region both for 20- and 41-volt photons, this method necessarily implies a large background signal. In addition, the metastable atoms 2^1S_0 of He I produced in the bombardment are able to liberate electrons from the photosensitive surface which are indistinguishable from the photoelectrons due to the 20-volt radiation. Originally it was hoped to distinguish between the 41-volt and 20-volt photons by a retarding potential placed between photocathode and anode, but too few of the photoelectrons had the theoretical maximum energy for this method to succeed.

Instead, the signal to background ratio was increased by inserting a film of collodion of a few hundred Ångström units thickness between the bombardment

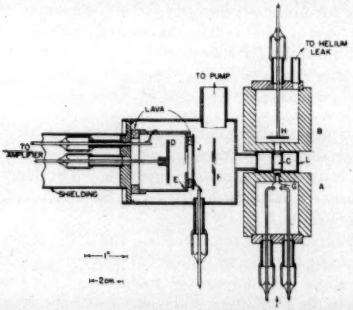


Figure 35. Cross section of apparatus used by Lamb and Skinner for determination of fine structure of singly ionized helium. The radiation of wavelength about 1·6 cm. was transmitted through the K-band waveguide C. This guide could be removed to permit the passage of longer wavelengths through the X-band guide L. The disc with a circular hole J defined the solid angle subtended by the detector plate at the excitation region. J was maintained at a positive potential with respect to D and E to keep electrons liberated from J away from D.

and detector regions. This completely removed the metastable atoms, and furthermore, preferentially absorbed the softer photons. According to studies by O'Bryan (1932) on a similar material a 100 A. thickness of celluloid transmits 60% of 600 A. radiation and 80% of 300 A. radiation.

Even with the help of the collodion film there was a background eight times as large as the signal due to the 41-volt radiation from the 2^2P states of the ion, but it was then possible to obtain significant resonance curves.

A cross-sectional diagram of the apparatus used by Lamb and Skinner is shown in Figure 35. A and B are pole pieces of a magnet, C is a waveguide and F is the collodion film. G is a hot filament maintained at a negative potential, so that electrons of several hundred volts energy stream across the waveguide into the opposite pole piece and are collected on a metal plate at H. The photons are detected with a photoelectric detector, which consists of a copper plate D and an electron collector E. The current from D to E is measured with an electrometer tube and a sensitive galvanometer.

As in the experiment with hydrogen the measurements are carried out in the presence of a magnetic field. The Zeeman splittings of the relevant levels are shown in Figure 36 and the transition frequencies to be expected as a function of magnetic field are shown in Figure 37.

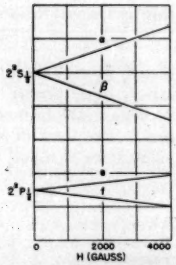


Figure 36. Zeeman splitting of $2^2S_{1/2}$ and $2^2P_{1/2}$ states of singly ionized helium.

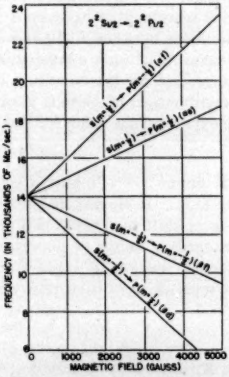


Figure 37. Zeeman transition frequencies as a function of the magnetic field for singly ionized helium.

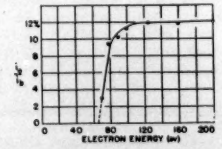


Figure 38. Radio-frequency quenching as a function of the bombarding electron energy. The ordinate is the percentage increase in the detector current when the radio-frequency power is turned on. This curve was taken with a thick collodion film in the apparatus, and with a large excess of radio-frequency power.

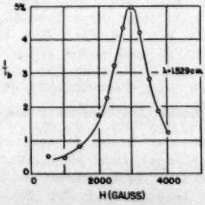


Figure 39. Experimental resonance curve for α transition in singly ionized helium.

8.4. *Results for He II Fine Structure*

Figure 38 shows the percentage increase in detector current produced by the introduction of a large amount of radio-frequency power of the resonance frequency as a function of the energy of the bombarding electrons. For electron energy above 100 ev. there is a 12% effect. There is no indication of quenching below 65 ev., the threshold for the production of metastable ions.

The resonance curves for the $2^2S_{1/2}$ $(m=\frac{1}{2}) \rightarrow 2^2P_{1/2}$ $(m=-\frac{1}{2})$ transition were obtained with a peak quenching of 4 or 5%. In obtaining resonance curves several fixed radio frequencies were used, resulting in quite widely differing

peak magnetic fields. The slope of the line obtained by plotting the frequency as a function of the peak magnetic field was equal to 1·87 Mc/s. per gauss, showing that the resonances did indeed belong to the αf transition. The sources of radio-frequency power were continuous wave magnetrons made especially for the experiment. The most accurate data were taken with radio-frequency wavelengths between 1·52 cm. and 1·61 cm., corresponding to peak magnetic fields between 3,050 and 2,450 gauss.

An example of such a resonance curve is shown in Figure 39. Only a very small correction for the curvature of the Zeeman frequency curves is necessary, or for asymmetries of various kinds.

The experimental value for the $2^2S_{1/2}$–$2^2P_{1/2}$ level shift in ionized helium was

$$\delta\nu_2 = (14{\cdot}02 \pm 0{\cdot}10) \times 10^3 \text{ Mc/s.} \qquad \ldots\ldots(68)$$

which is about 1·4% higher than the present calculated value according to equation (15). It should be possible to obtain a relatively more precise value for the level shift in ionized helium than for hydrogen, since in the former case the complicating factor of hyperfine structure is missing. The work of Lamb and Skinner was of an exploratory nature, and apparatus for a much more accurate measurement is currently being built.

REFERENCES

Anderson, C. D., 1932, *Science*, **76**, 238.
Bechert, K., and Meixner, J., 1935, *Ann. Phys., Lpz.*, **22**, 525.
Bethe, H. A., 1933, *Handb. der Phys.*, **24** (1), 2nd edn. (Berlin : Springer); 1947, *Phys. Rev.*, **72**, 339.
Bethe, H. A., Brown, L. M., and Stehn, J. R., 1950, *Phys. Rev.*, **77**, 370.
Betz, O., 1932, *Ann. Phys., Lpz.*, **15**, 321.
Blatt, J. M., 1945, *Phys. Rev.*, **67**, 205.
Bleakney, W., 1930, *Phys. Rev.*, **35**, 1180.
Bleakney, W., and Smith, L. G., 1936, *Phys. Rev.*, **49**, 402.
Bonhoeffer, K. F., 1927, *Ergebn. Exakt. Naturw.*, **6**, 201.
Breit, G., 1947, *Phys. Rev.*, **72**, 984.
Breit, G., and Brown, G. E., 1948, *Phys. Rev.*, **74**, 1278.
Breit, G., and Rabi, I. I., 1931, *Phys. Rev.*, **38**, 2082.
Breit, G., and Teller, E., 1940, *Astrophys. J.*, **91**, 215.
Buehl, A., 1933, *Helv. phys. Acta*, **6**, 231.
Caldirola, P., 1948, *Nuovo Cim.*, **5**, 207.
Chu, D. Y., 1939, *Phys. Rev.*, **55**, 423.
Cobas, A., and Lamb, W. E., Jr., 1944, *Phys. Rev.*, **65**, 327.
Compton, K. T., 1922, *J. Opt. Soc. Amer.*, **6**, 910.
Compton, K. T., and Russell, H. N., 1924, *Nature, Lond.*, **114**, 86.
Coolidge, A. S., James, H. M., and Present, R. D., 1938, *J. Chem. Phys.*, **6**, 730.
David, E., 1949, *Z. Phys.*, **125**, 274.
Davis, L. Jr., Feld, B. T., Zabel, C. W., and Zacharias, J. R., 1949, *Phys. Rev.*, **76**, 1076.
Dirac, P. A. M., 1928, *Proc. Roy. Soc.* A, **117**, 610; 1931, *Ibid.*, **133**, 60.
Drinkwater, J. W., Richardson, O., and Williams, W. E., 1940, *Proc. Roy. Soc.* A, **174**, 164.
Dorrestein, R., 1942, *Physica*, **9**, 433, 447.
Duffendack, O. S., 1922, *Phys. Rev.*, **20**, 655.
Dyson, F. J., 1949, *Phys. Rev.*, **75**, 486.
Feynman, R. P., 1949, *Phys. Rev.*, **76**, 769.
Fowles, G. R., 1948, *Phys. Rev.*, **74**, 219.
Franck, J., and Jordan, P., 1926, *Anregung von Quantensprüngen durch Stösse* (Berlin : Springer), p. 117.
French, J. B., and Weisskopf, V. F., 1949, *Phys. Rev.*, **75**, 1240.

FRÖHLICH, H., HEITLER, W., and KAHN, B., 1939 a, *Proc. Roy. Soc.* A, **171,** 269; 1939 b, *Phys. Rev.*, **56,** 961.
GIULOTTO, L., 1947 a, *Ric. Sci.*, **17,** No. 2–3; 1947 b, *Phys. Rev.*, **71,** 562.
GORDY, W., 1948, *Rev. Mod. Phys.*, **20,** 668 (review article).
DE GROOT, W., and PENNING, F. N., 1933, *Handb. der Phys.*, **23** (1), 2nd edn. (Berlin: Springer), p. 78.
HAASE, T., 1935, *Ann. Phys., Lpz.*, **23,** 675.
HIRSCHBERG, J. G., Jr., and MACK, J. E., 1950, *Phys. Rev.*, **77,** 745.
HOUSTON, W. V., 1937, *Phys. Rev.*, **51,** 446.
HOUSTON, W. V., and HSIEH, Y. M., 1934, *Phys. Rev.*, **45,** 263.
HYLLERAAS, E. A., 1931, *Z. Phys.*, **71,** 739.
JAUCH, J. M., 1940, *Helv. Phys. Acta*, **13,** 451.
KAHN, B., 1941, *Physica*, **8,** 58.
KELLOGG, J. M. B., and MILLMAN, S., 1946, *Rev. Mod. Phys.*, **18,** 323 (review article).
KEMBLE, E. C., and PRESENT, R. D., 1933, *Phys. Rev.*, **44,** 1031.
KOPFERMANN, H., KRÜGER, H., and ÖHLMANN, H., 1949, *Z. Phys.*, **126,** 760.
KOPFERMANN, H., and PAUL, W., 1948, *Nature, Lond.*, **162,** 33.
KROLL, N. M., and LAMB, W. E., Jr., 1949, *Phys. Rev.*, **75,** 388.
KUHN, H., and SERIES, G. W., 1948, *Nature, Lond.*, **162,** 373.
KUSCH, P., and FOLEY, H., 1948, *Phys. Rev.*, **74,** 250.
LAMB, W. E., Jr., 1939, *Phys. Rev.*, **56,** 384; 1940, *Ibid.*, **57,** 458.
LAMB, W. E., Jr., and RETHERFORD, R. C., 1947, *Phys. Rev.*, **72,** 241; 1949 a, *Ibid.*, **75,** 1325; 1949 b, *Ibid.*, **75,** 1332; 1950, *Ibid.*, **79,** 549.
LAMB, W. E., Jr., and SKINNER, M., 1950, *Phys. Rev.*, **78,** 539.
LEO, W., 1926, *Ann. Phys., Lpz.*, **81,** 757; 1937, *Phys. Z.*, **38,** 921.
MACK, J. E., and AUSTERN, N., 1947, *Phys. Rev.*, **72,** 972.
MACK, J. E., and BARKOFSKY, E. C., 1942, *Rev. Mod. Phys.*, **14,** 82.
MASSEY, H. S. W., 1930, *Proc. Camb. Phil. Soc.*, **26,** 386; 1931, *Ibid.*, **27,** 460.
MICHELSON, A. A., and MORLEY, E. W., 1887, *Phil. Mag.*, **24,** 46.
MURAKAWA, K., SUWA, S., and KAMEI, T., 1949, *Phys. Rev.*, **76,** 1721.
NAFE, J. E., and NELSON, E. B., 1948, *Phys. Rev.*, **73,** 718.
O'BRYAN, H. M., 1932, *J. Opt. Soc. Amer.*, **22,** 739.
OLIPHANT, M. L. E., 1929, *Proc. Roy. Soc.* A, **124,** 228.
OLMSTEAD, P. S., and COMPTON, K. T., 1923, *Phys. Rev.*, **22,** 559.
OPPENHEIMER, J. R., 1930, *Phys. Rev.*, **35,** 461.
PASCHEN, F., 1916, *Ann. Phys., Lpz.*, **50,** 901; 1927, *Ibid.*, **82,** 689.
PASTERNACK, S., 1938, *Phys. Rev.*, **54,** 1113.
PRESENT, R. D., 1935, *J. Chem. Phys.*, **3,** 122.
ROJANSKY, V., 1929, *Phys. Rev.*, **33,** 1.
ROJANSKY, V., and VAN VLECK, J. H., 1928, *Phys. Rev.*, **32,** 327.
SCHWINGER, J., 1948 a, *Phys. Rev.*, **73,** 416; 1948 b, *Ibid.*, **74,** 1439; 1949 a, *Ibid.*, **75,** 651; 1949 b, *Ibid.*, **76,** 790.
SLOTNICK, M., and HEITLER, W., 1949, *Phys. Rev.*, **75,** 1645.
SOMMERFELD, A., 1916, *Ann. Phys., Lpz.*, **51,** 1; 1941 a, *Naturwissenschaften*, **29,** 286; 1941 b, *Z. Phys.*, **118,** 295.
SOMMERFELD, A., and UNSÖLD, A., 1926 a, *Z. Phys.*, **36,** 259; 1926 b, *Ibid.*, **38,** 237.
SONKIN, S., 1933, *Phys. Rev.*, **43,** 788.
TOMONAGA, S., 1946, *Prog. Theor. Phys.*, **1,** 27, and subsequent papers in that journal.
UEHLING, E. A., 1935, *Phys. Rev.*, **48,** 55.
WEBB, H. W., 1924, *Phys. Rev.*, **24,** 113.
WEISSKOPF, V. F., 1934, *Z. Phys.*, **90,** 817; 1939, *Phys. Rev.*, **56,** 72.
WELTON, T. A., 1948, *Phys. Rev.*, **74,** 1157.
WHITE, H. E., 1934, *Introduction to Atomic Spectra* (New York: McGraw-Hill).
WILLIAMS, R. C., 1938, *Phys. Rev.*, **54,** 558, gives references to previous work. The studies before 1930 are mostly listed by F. K. RICHTMYER, *Introduction to Modern Physics*, 1934, 2nd edn. (New York: McGraw-Hill), p. 398.
WOOD, R. W., 1921, *Phil. Mag.*, **42,** 729.
WOOLEY, H. W., SCOTT, R. B., and BRICKWEDDE, F. G., 1948, *Bur. Stand. J. Res., Wash.*, **41,** 379.

NEW TECHNIQUES IN OPTICAL INTERFEROMETRY

By H. KUHN
The Clarendon Laboratory, Oxford

CONTENTS

PART I: THIN FILMS AND INTERFERENCE FILTERS

PAGE

§ 1. Introduction 65
§ 2. Dielectric films 65
§ 3. Metal films 68
§ 4. Dielectric films on metal 69
§ 5. The intensity distribution in multiple beam fringes 70
§ 6. Interference filters 71
§ 7. Polarization filters 74

PART II: SPECTROSCOPIC INTERFEROMETRY

§ 8. The single etalon 74
§ 9. The double etalon 77
§ 10. Photoelectric recording 79
§ 11. Solar interferometry 80

PART III: STANDARDS AND TESTING METHODS

§ 12. Light sources and wavelength standards 81
§ 13. Gauge testing methods 83
§ 14. Testing of wavefronts 83
§ 15. Testing of interferometer plates 84

PART IV: THE STUDY OF SURFACE STRUCTURES

§ 16. The principle of the method 85
§ 17. Fringes in monochromatic light 86
§ 18. Heterochromatic fringes 87
§ 19. Technical details and applications 88

PART V: INTERFERENCE REFRACTOMETERS AND INTERFERENCE MICROSCOPES

§ 20. Refractometry of homogeneous media 88
§ 21. Applications to aerodynamics 89
§ 22. Interference microscopes 91

ABSTRACT. The first section deals with the optical properties of thin films, especially their reflection and transmission coefficients. Multiple dielectric films, deposited on glass surfaces, can be made to have reflection coefficients up to 0·9 for visible light, with very small absorption losses. Considerably higher reflectivities can be reached with silver films, but the absorption losses are higher and increase with decreasing wavelength. Metallic reflectivity can be increased if one or two dielectric films of appropriate thickness are deposited on the metal surface. The intensity distribution in the reflected and transmitted light in multiple beam interference is discussed and applied to interference filters. 'Frustrated total reflection' filters and polarization filters are briefly described.

Part II deals with some new developments in the methods of using Fabry–Perot etalons. The resolving power of these instruments is, in practice, limited by geometrical imperfections of the glass or silica surfaces. The theory and practical performance of the double etalon, both with different and with equal spacings, the photoelectric method of recording etalon fringes and the use of etalons in solar spectroscopy are described.

Part III deals with standards and testing methods, pure isotope light sources, testing of gauges and of optical instruments, and gives a description of the 'wavefront shearing interferometer'.

The interferometric study of surfaces forms the subject of Part IV, especially the method of using multiple beam interference fringes. The various forms of the method, using Fizeau fringes and heterochromatic fringes, either in transmission or in reflection, are described and their applications briefly indicated.

In the last section some recent modifications of the classical methods of interference refractometry are discussed, and also applications of interferometers of the Mach–Zehnder type to the study of gas flow at high speeds.

The method of interference microscopy is outlined and some recent designs of interference microscopes are described.

I. THIN FILMS AND INTERFERENCE FILTERS

§1. INTRODUCTION

In many interferometers and other optical devices, surfaces are required to divide the incident light wave into transmitted and reflected waves whose amplitudes often have to bear a definite ratio which varies greatly according to the particular purpose. Such surfaces can be realized by the use of thin films of metal, by single or multiple films of dielectrics, or by combinations of both. The first section of this Report will be given to a review of methods which have been used, and which are being further developed, in producing these surfaces. The interference phenomena determining their optical properties will be briefly outlined; assemblies of films acting as 'interference filters' are of particular interest on account of their close connection with the Fabry–Perot etalon and other multiple interference devices.

In two-beam interferometry, the 'beam splitter' is generally required to produce transmitted and reflected beams of approximately equal intensity. This can be achieved by means of very thin metal films, or by single or multiple dielectric films.

Multiple beam interferences require surfaces of high reflectivity. If R and T are the reflection and transmission coefficients, defined by the ratio of the intensity of the reflected or transmitted beam to that of the incident beam, the absorption defined by $A = 1 - R - T$ has to be as low as possible. This can be achieved by simple metal films of appropriate thickness, by metal films covered with dielectric films, or by multiple dielectric films of alternately high and low refractive index. The reflectivity of the latter depends critically on the wavelength.

A different application of interferences in single and double dielectric films is made in the anti-reflection coatings which are now frequently used in optical instruments.

Polarization filters are based on interferences of a type different from that in interference filters, but on account of the similar effect which they achieve they are included in this section.

§2. DIELECTRIC FILMS

The reflectivity of a surface between two media of refractive indices n_1 and n_2 for normally incident light is given by Fresnel's formulae as $R = \{(n_1 - n_2)/(n_1 + n_2)\}^2$. For an air–glass boundary the value of this expression is about 0·04. Transparent dielectrics with refractive indices over 2·5 for visible light are scarcely available, so that the reflectivity of a single surface cannot be made much greater than 0·15.

The modern technique of high vacuum evaporation has made it possible to produce very thin uniform dielectric films whose thickness can be accurately controlled. The following substances are generally used:

Low refractive index: cryolite $n=1{\cdot}35$; MgF_2 $n=1{\cdot}38$

High refractive index: ZnS $n=2{\cdot}3$; TiO_2 $n=2{\cdot}45$; PbS_2 and Sb_2S_3

If a film of low refractive index is deposited on glass and the optical thickness of the film made equal to $\lambda/4$, where λ is the wavelength of the light *in vacuo*, the waves reflected normally from the two surfaces have a phase difference of π. If, further, the refractive index of the film is the geometric mean of the indices of the supporting base and air, the two waves have equal amplitude and cancel one another. The dielectric then acts as an 'anti-reflection film' (Strong 1936, Blodgett 1939, 1940, Cartwright and Turner 1939 a, Jacquinot 1942, Greenland 1943). Magnesium fluoride and cryolite approach these conditions fairly well. The exact value of the refractive index of vacuum-deposited films is generally lower than that of the bulk material and depends on the conditions of deposition (Perrot 1949, Rood 1949). The latter author even reports ageing effects resulting in change of refractive index.

If a film of high refractive index is deposited on glass, and the optical thickness again chosen as $\lambda/4$, the waves reflected from the two surfaces are in phase (one of them having suffered a phase change of π on external reflection), and their interference leads to strongly enhanced reflectivity of the surface. If multiple reflections in the film are taken into account, the reflection coefficient is found to be (Messner 1943)

$$R=(n_2^2-n_1n_3)^2/(n_2^2+n_1n_3)^2,$$

where n_1, n_2 and n_3 are the refractive indices of air, film and support. The value of R increases with increasing n_2 and reaches the value 0·38 for $n_2=2{\cdot}5$, with $n_1=1$ and $n_3=1{\cdot}5$.

Multiple films of dielectrics allow a greater variety of effects to be achieved. Anti-reflection films which are highly effective over a wide range of wavelengths can be made by depositing several layers whose indices decrease progressively from the support towards the outside (King and Lockhart 1946, Lockhart and King 1947, Hiesinger 1948). Dielectric films arranged in this order can, in fact, only decrease and never increase the reflectivity of the surface, whatever their thickness may be.

If a number of films of alternately high and low refractive index are deposited on glass, the reflectivity can rise to very high values. Cartwright and Turner (1939 b), using films of TiO_2 and cryolite, claim to have reached a reflectivity of 0·85. Messner (1943) reports reflectivities of 0·3, 0·6 and 0·8 for white light, with the use of one, three and five films respectively. Dufour (1948), using ZnS and cryolite, finds values of 0·6, 0·8 and 0·9 for three, five and seven layers (Figure 1 (*a*)). The latter value appears to be the present upper limit achieved by this method. Greenland (1946) also describes the method.

With increasing number of layers the reflectivity depends on the wavelength to an increasing extent. When viewed in reflected and transmitted light under

the same angle, such filters show almost exactly complementary colours, owing to a bright band in the reflected spectrum and a dark band in the transmitted spectrum. The relation $R+T=1$ holds to a very high degree of approximation for the greater part of the visible spectrum. Only violet light tends to be absorbed noticeably. For ultra-violet light no reflectors of this kind appear to have been made.

The optical properties of a dielectric medium of periodically varying dielectric constant have long been known in the phenomenon of iridescence of crystals caused by the appearance of periodically repeating twinning layers, e.g. in potassium chlorate (see, for example, R. W. Wood 1936). Lord Rayleigh (1888) first explained this phenomenon, and more recently, a large number of authors have treated the theory of the propagation of light through multiple layers, some including the more difficult case of conducting layers (Subrahmanian 1941, Ramachandran 1942, Mooney 1945, 1946, Caballero 1947, Weinstein 1947, Vasicek 1947, 1949, Crook 1948, Abelès 1948, 1949, Hiesinger 1948, Polster 1949). Discrepancies between the results of Vasicek and Crook were found to be due to an error in one formula of the first author (Wilcock 1949).

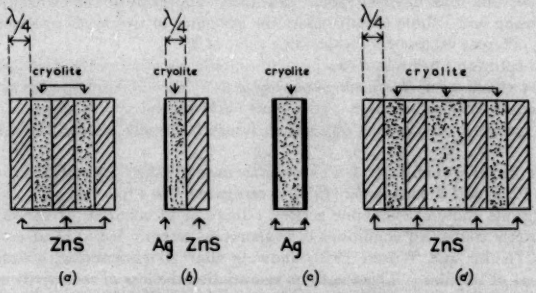


Figure 1. Applications of thin films.

(*a*) Multiple dielectric reflector.
(*b*) Increase of metallic reflection.
(*c*) Interference filter with metal films.
(*d*) Interference filter with dielectric films.

There are essentially two methods for calculating the optical properties of multiple layers. Either Fresnel's formulae are applied, and the (complex) amplitudes of all possible refracted and reflected waves are added in each layer. This leads to sums of geometrical progressions which are easy to evaluate for a small number of layers. An alternative method which appears preferable in more complicated cases, starts directly from Maxwell's equations and the boundary conditions. Born (1933) illustrates both methods by applying them to a single film and deriving Airy's formula. Developing the latter method further, Kronig, Blaisse and Sande (1938) and Muchmore (1948) apply methods of network analysis and the concept of impedance.

By any of these methods, the calculations tend to be cumbersome, and graphical methods such as those developed by Perrot (1943), Malé (1945) and Cotton (1947) might prove to be of great value.

§3. METAL FILMS

Silver films are still the means most frequently used to obtain high reflectivities for visible light. If greater absorption losses can be tolerated but higher resistance to chemical attack is required, aluminium films are used, the latter also as reflectors for ultra-violet light. Rhodium films offer the advantage of very much greater chemical and mechanical stability at the expense of still greater absorption losses.

The method of chemical silvering has been entirely superseded by evaporation in a high vacuum. Cathodic sputtering does not appear to offer any advantages and seems to be going out of use, though it has not been proved that sputtered films are optically inferior to vacuum deposited films.

Extremely thin, i.e. highly transparent, films of silver and other metals show some interesting optical properties but are not used much in interferometry. Rouard (1932, 1949) found that the reflectivity, for light incident from the glass side, when plotted as a function of film thickness, passes through a minimum value which is nearly equal to zero for the whole of the visible spectral range, so that internal reflection of white light can be almost entirely suppressed. Such extremely thin films of silver (about 25 A. thick) appear to be somewhat unstable, and Strong and Dibble (1940) report the existence of two types of silver films, having different values of R for the same value of T.

For splitting a beam into two parts of roughly equal intensity, metal films are not very effective; if R is made about equal to T, then absorption A is generally about equal to either of these. It appears likely that silver and aluminium films which are often used in this way will in future be largely replaced by dielectric films.

For thicker silver films ($R \gg T$), such as are used in Fabry–Perot interferometers, Goos (1936) and Krautkrämer (1938) investigated R as a function of T and λ, but their results show considerable scatter. Increase of accuracy of measurement and strictly controlled conditions of evaporation recently led to more consistent results (Kuhn and Wilson 1950) showing that A increases with decreasing thickness of the film. These authors also studied the loss of reflectivity of silver films as a result of ageing. Figure 2 shows $R+T$ plotted as a function of T. The dotted curves refer to films which had been kept for three weeks in only moderately clean air. Sennett and Scott (1950) recently found a very marked influence of the rate of deposition on the structure and optical properties of silver films. Slow deposition leads to films showing agglomeration structure and high absorption losses.

For ultra-violet light, aluminium films are generally used for producing high reflectivity, either in opaque or transparent films, but few quantitative data exist. Their properties depend on vacuum conditions and other factors much more critically than those of silver films. Crawford, Gray, Schawlow and Kelly (1949) find that absorption losses increase with decreasing rate of evaporation. It appears that, throughout the visible and ultra-violet range, A is hardly ever less than 0·1 and often much more.

Even less is known of the reflectivity of other metals in the ultra-violet. Preliminary reports of very high values of R for aluminium–magnesium alloys (Edwards 1933) still lack confirmation.

The theoretical treatment of the optical properties of thin metal films is similar to that of dielectrics. It is included in some of the papers quoted in that

connection, and Laurin (1906), Barns and Czerny (1931), Murmann (1933), Wolter (1937) and Hadley and Dennison (1946, 1947, 1948) made further contributions. The fact that the refractive index is complex, containing two constants, makes the calculations very complicated. They have been largely confined to the infra-red range where the real and imaginary parts of the refractive index can be assumed equal; this simplifies the calculations somewhat. The exact value of the thickness of a transparent metal film is extremely difficult to determine, and the values of the optical constants depend not only on the thickness but on the conditions of deposition of the film. These facts severely restrict the practical uses of theoretical treatments.

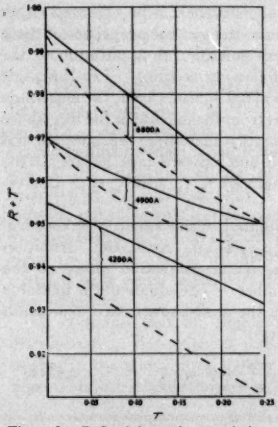


Figure 2. Reflectivity and transmission of silver films.
——— new films – – – – aged films.

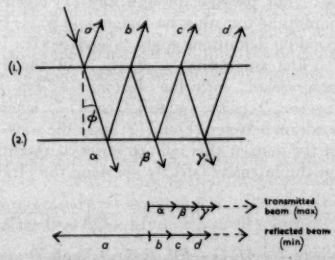


Figure 3. Amplitudes of transmitted and reflected beams.

For the purposes of practical interferometry, one generally regards a metal film as defined by the following properties which, if required, have to be experimentally determined for each individual film: the transmission factor T, the external reflectivity R, the reflectivity from the glass–metal surface R' and the phase changes which a wave suffers on either of these reflections and on transmission. For fairly dense films, R' is always slightly smaller than R for visible light.

§4. DIELECTRIC FILMS ON METAL

A dielectric film is sometimes used for protecting a metal surface from corrosion. If the thickness of the dielectric layer is so chosen that the waves reflected from the two surfaces are in phase, the reflection coefficient of the metal is slightly increased. A much greater increase can be produced by covering the metal surface with one film of low followed by one of high refractive index (Turner 1946, Messner 1947, Hammer and Miller 1949) (see Figure 1(*b*)). These

authors report an increase of the reflectivity of a silver film, for white light, from 0·98 to 0·995, and of an aluminium film from 0·91 to 0·96 and, in another case, to 0·98.

The absorption is reduced because part of the reflection is effected by the dielectric film which is free from losses. A given amplitude of the resultant reflected beam is produced by a smaller amplitude of the beam reflected from the metal surface, and it is the latter on which the losses depend.

§ 5. THE INTENSITY DISTRIBUTION IN MULTIPLE BEAM FRINGES

For two partially transparent metal films, forming the boundaries of three dielectric media (see Figure 3), interference phenomena can be observed either in transmitted or reflected light. In this section, the general properties of these interferences are to be summarized in a form suitable for application to the various phenomena treated in this and the following sections. Let the angle formed by the normal and the ray between the metal surfaces be ϕ, the amplitude reflection factors of surface (1) for light incident from the outside or the interspace r_1 and r_1' respectively, and the phase changes occurring in these reflections ρ_1 and ρ_1'. For passage through film (1) downward and upward (see Figure 3) the amplitude transmission factors and phase changes are t_1, t_1' and τ_1, τ_1'. Corresponding definitions apply to surface (2).

The amplitudes α, β, γ.... of the transmitted beams form a geometrical progression with the factor $r_1'r_2' \exp\{i(\rho_1' + \rho_2' + \Delta)\}$, with the definition $\Delta = 4\pi n_2 D \cos\phi \times \nu = (4\pi n_2 D \cos\phi)/\lambda$, where n_2 is the refractive index of the medium between (1) and (2) and ν the wave-number. The square of the modulus of the sum of this infinite series is essentially the well-known Airy expression for the intensity (strictly speaking the brightness)

$$I(\Delta) = \left(\frac{T}{1-R}\right)^2 \times \frac{I_0}{1 + F\sin^2(\rho' + \rho_2' + \Delta)/2} \qquad \ldots\ldots(1)$$

where $F = 4R/(1-R)^2$, $R = r_1'r_2'$ and $T = t_1t_2$.

The function $I(\Delta)$ will be referred to as 'intensity distribution'. For colour filters, the variation of Δ is produced essentially by the factor $1/\lambda$, for monochromatic fringes of the Fabry–Perot type by the factor $\cos\phi$, and for Fizeau fringes by the factor D. The intensity distribution in the transmitted system, contained in the formula (1) is well known, and only a few points need be mentioned here. The peaks are symmetrical, and they are the narrower the greater the value of R. The phase changes ρ_1' and ρ_2', for small values of ϕ, merely have the effect of changing the 'effective' value of the spacing D, but do not influence the intensity distribution.

If we define the half-value width $\Delta_{1/2}$ as the difference of two adjacent values of Δ for which $I = I_{max}/2$, then $\Delta_{1/2}/2\pi$ is the half-value width expressed as a fraction of one order. The reciprocal of this value, $2\pi/\Delta_{1/2}$, may be called N and can be interpreted as 'effective number of beams' (see § 8). If N is sufficiently large, the sine in (1) can be replaced by the arc, and we find after substitution of the value of F,

$$N = \pi\sqrt{R}/(1-R). \qquad \ldots\ldots(1a)$$

The properties of the reflected beam in multiple reflection interferences are much more complicated. Hamy (1906) studied these interference phenomena

in reflection very thoroughly, both theoretically and experimentally, for chemically deposited silver films, but these fringes received little attention until recently (Tolansky 1948). Holden (1949) repeated Hamy's experiments for vacuum deposited films, mainly with a view to applications to surface topology.

The peculiar properties of the reflected system are due to the fact that the first beam *a* (Figure 2) has a phase difference compared with *b* altogether different from that between any other two successive beams. Summation of the amplitudes *b*, *c*.... leads to the same expression as that of the transmitted beams α, β..... For a value of Δ for which the transmitted light has maximum intensity, the sum of *b*, *c*.... therefore also has a maximum, i.e. all the amplitudes *b*, *c*.... are in phase. The condition for this is $\rho_1'+\rho_2'+\Delta=n2\pi$ (where n is an integer). The phase difference between *a* and *b* is $\delta_{a,b}=\tau_1+\tau_1'+\rho_2'-\rho_1+\Delta$, for which we find, by substituting the above condition and ignoring multiples of 2π,

$$\delta_{a,b}=\tau_1+\tau_1'-\rho_1-\rho_1'. \qquad \ldots\ldots(2)$$

If (1) and (2) are unsilvered surfaces of dielectrics, such as the surfaces of a glass plate, either ρ_1 or ρ_1' has the value π, while the other three terms in (2) vanish, so that $\delta_{a,b}=\pi$. Airy has already shown for this case that the reflected system is exactly complementary to the transmitted system, i.e. that for any given angle the sum of the intensities of the transmitted and reflected system is equal to the intensity of the incident light. It is obvious that this must be true for any partially reflecting surfaces for which no energy is lost ($A=0$), since the radiation incident within any given element of solid angle remains within the corresponding elements of solid angle in the transmitted and reflected waves. If the two surfaces have equal reflectivity, then the minima in the reflected system are of zero intensity.

It is found that for highly reflecting silver films ($R>0{\cdot}8$), the intensity distribution in the reflected interference system is very nearly complementary to that in the transmitted system, forming narrow, symmetrical minima. One must therefore assume that for such nearly ideal films (films with small values of A)

$$\tau_1+\tau_1'-\rho_1-\rho_1'=(2n+1)\pi.$$

The summation of the amplitudes in this case is shown in the vector diagram Figure 3.

The manner in which the phase change is distributed among the four terms of equation (2) does not seem to be known, but in the limit of high reflectivity films, both ρ_1 and ρ_1' approach the value π; this would indicate a phase change of $\pi/2$ for each transit.

Hamy has shown that for lower reflectivities ($R<0{\cdot}8$) the peaks in the reflected system are asymmetrical, but that they become symmetrical again for very low reflectivities ($R<0{\cdot}3$), with the maxima in the positions of the minima of the reflection fringes found for high values of R. The intensity of these fringes can, under favourable conditions, approach the intensity of the incident light fairly closely, but the width of the peaks is rather large, about one-fifth of one order.

§6. INTERFERENCE FILTERS

The principle of the interference filter is the same as that underlying the formation of the heterochromatic fringes known as Edser–Butler fringes. As the air film between the partially reflecting surfaces is reduced in the Edser–Butler experiment, the fringes become more widely spaced in the spectrum until only a

small number of transmission bands, possibly only one, remains; the double film then acts as an interference filter. The technique of depositing thin films of dielectrics by evaporation *in vacuo* provided the practical means of producing such filters.

One film of silver, one film of a substance such as cryolite or magnesium fluoride and a second film of silver, approximately equal in thickness to the first, are successively deposited on a glass plate (Figure 1 (*c*)). The thickness of the dielectric film determines the wavelength of the bands; it is generally controlled optically during the evaporation (Dufour 1948, Greenland and Billington 1950).

It appears that interference filters were first developed in Germany during the last war (Geffcken 1942, 1948, 1949)*, but they soon came to be used widely, especially as infra-red filters, and can now be obtained commercially. Hadley and Dennison (1946, 1947, 1948) give a detailed discussion of the theory and a number of experimental facts, with the emphasis on the infra-red spectrum.

The performance of an interference filter can be described in the main by four factors:

(i) The bandwidth: this can be defined as the half-value width, i.e. the difference $\Delta\nu_{1/2}$ or $\Delta\lambda_{1/2}$ between the two wave-numbers or wavelengths for which the transmission has half the maximum value. It is determined by the order of interference n and by $R=r_1'r_2'$, with the definitions introduced above. R is the geometrical mean of the internal reflectivities of the two surfaces. From (1) and (1 *a*) and from the relation $\partial\nu/\nu=\partial\Delta/\Delta=\partial\Delta/2\pi n$ the following approximate formula can easily be derived:

$$\Delta\lambda_{1/2}/\lambda\simeq\Delta\nu_{1/2}/\nu=1/nN=(1-R)/n\pi\sqrt{R}. \qquad \ldots\ldots(3)$$

(ii) The maximum transmission factor, which can be read from (1) as

$$(I/I_0)_{\max}=[T/(1-R)]^2. \qquad \ldots\ldots(4)$$

(iii) The 'contrast factor', defined as ratio of maximum to minimum transmission factor,

$$I_{\max}/I_{\min}=\{(1+R)/(1-R)\}^2. \qquad \ldots\ldots(5)$$

(iv) The maximum permissible angular divergence of the incident light. Expansion of the term $\cos\phi$ in the interference condition gives the change of the angle ϕ' of the rays in air which causes a shift $\Delta\lambda$ of the wavelength band,

$$\phi'=n_2(2\Delta\lambda/\lambda)^{1/2}$$

(for small ϕ) where n_2 is the refractive index of the dielectric.

The thicker the silver films are, the smaller is the band width and the greater the contrast factor, but also the smaller the transmission factor. The relation between these three depends on the value of R which can be achieved for a given T and a given wavelength. Experimental data for the reflectivity of silver–dielectric surfaces appear to be very scanty. As a rough guide, the data given in §3 for reflection from silver–air surfaces can be used, with slightly lower values of R.

The orders of magnitude may be illustrated by a numerical example corresponding roughly to what can be achieved in practice. Assuming $R=0{\cdot}9$, $T=0{\cdot}05$, $\lambda=6000$ A., the value of $\Delta\lambda$ is found to be $200/n$ A. For a thickness of the dielectric producing a second order filter, the bandwidth would be 100 A.

* See also Tolansky (1948).

The only other band in the visible spectrum would be at $\lambda = 4000$ A. ($n = 3$) which could be filtered out by means of an ordinary dye filter with little loss of intensity. The transmission factor is found to be $\frac{1}{4}$, the contrast factor 360. Regarding the latter, however, the assumption of the same value of R for the wavelength of the minimum is very unrealistic. In order to find the permissible divergence of the light, one may assume half the value of $\Delta\lambda_{1/2}$ as the permitted variation of wavelength. One then finds, for $n_2 = 1{\cdot}4$, that ϕ_{max} is 0·17. Variation of the phase change ρ' with the angle of incidence is disregarded.

All these estimates neglect imperfections in the surfaces owing to which the actual performance will fall somewhat below the quoted values. Nevertheless, these orders of magnitude show that interference filters compare very favourably with ordinary colour filters using dyes and are likely to replace them in future for many purposes.

The spectrum of white light reflected from an interference filter is roughly complementary to the spectrum of the transmitted light and shows narrow minima of reflection which are hardly of practical value. If, however, one film is made highly reflecting and the other very thin, the intensity distribution is that of the low reflectivity type fringes described by Hamy. Reflection filters of this kind have been used for centimetre waves and for infra-red light (Hadley and Dennison 1946, 1947, 1948). If the thickness of the transparent metal film is suitably chosen the transmission in the minima can be made practically zero.

Dufour at a Conference at Marseille in 1949 discussed the construction of filters containing three metal films separated by dielectric layers; such filters should show very great contrast factors.

As mentioned in § 2, interference filters of the Fabry–Perot type can be made from dielectric films only (Mooney 1945, 1946, Banning 1947). Each metal film is replaced by a multiple dielectric reflector, and the spacer consists of another dielectric film, usually of low refractive index. These filters have a maximum transmission of almost 1, but their bandwidth is not very small (Figure 1(*d*)).

The increase of metallic reflection by dielectric layers can, of course, also be used in the construction of interference filters. In an arrangement of this kind, films of cryolite and ZnS adjacent to the silver films increase the reflectivity of the latter, while a cryolite film in the middle acts as spacer.

If two media of high refractive index are separated by a very thin film of a medium of low refractive index, light incident at an angle greater than the critical angle is not totally reflected. By suitable choice of the thickness of the film, one can thus produce very highly reflecting and partially transmitting surfaces which are almost free from absorption. This fact, which has long been known, was recently used for the design of an interference filter known as 'frustrated total reflection' filter (Leurgans and Turner 1947). Figure 4 gives a diagram of this device. The thickness of the layer of low refractive index determines the reflectivity and thus the bandwidth, while the thickness of the layer of high refractive index determines the wavelength of the transmitted band. Extremely narrow transmission bands can be produced if the angle of incidence is kept within small limits. Billings and Pittman (1949) describe an infra-red filter of this type, made of NaCl prisms and vacuum deposited films of NaF and AgCl. Gee and Polster (1949) use the same principle as a highly sensitive means of studying very small imperfections in thin films of dielectrics.

§7. POLARIZATION FILTERS

Interference phenomena in polarized light can also be used for the construction of colour filters. If a plate of birefringent material, with the optic axis in the plane of the plate, is placed between two polarizers whose optic axes are parallel to one another and at an angle of 45° to the axis of the plate, normally incident light is transmitted if the difference of the optical paths of ordinary and extraordinary rays is an integral multiple of λ. If white light passes through this arrangement, a spectrograph reveals a succession of bright and dark bands. The intensity distribution is a $\cos^2$ function, so that the transmission is zero in the minima.

Though the width of the bands can be made very small by the use of great optical paths, the closeness of the bands makes the device, in this simple form, of only very limited practical use. R. W. Wood applied it, for example, to filter out one of the two components of the sodium yellow doublet. Lyot (1933, 1944) has developed the method and has succeeded in constructing a powerful colour filter which he and others applied to investigations of the solar corona.

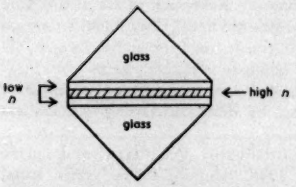


Figure 4. Frustrated total reflection filter.

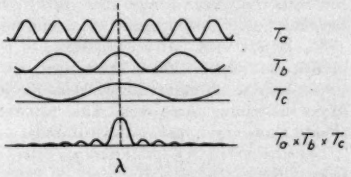


Figure 5. Mode of action of the Lyot filter.

Lyot placed a number of quartz or calcite plates, whose thicknesses were in the ratio 1:2:4:...., in series, each between two polarizers (Nicol prisms, polaroid films or Rochon prisms). The mode of action is illustrated by Figure 5 where T_a, T_b and T_c are the transmission curves for three such elements used individually and the curve at the bottom represents diagrammatically the transmission curve of the assembly which is the product of the three curves *a*, *b* and *c*.

Lyot has, in this way, built up filters containing six units, with a bandwidth of 3 A. and a transmission factor of 0·25; the filters are highly sensitive to temperature changes and have to be mounted in thermostats. Lyot used quartz wedges for adjusting the optical paths, Billings (1947) achieves this adjustment by using plastic sheets which are strained mechanically or by electro-optic effects (Kerr effect) for which he finds crystals of $NH_4H_2PO_4$ especially suitable.

II. SPECTROSCOPIC INTERFEROMETRY

§8. THE SINGLE ETALON

Among the three instruments of high resolving power, the echelon grating, the Lummer plate and the etalon, the latter has steadily gained in importance, owing to its flexibility in the choice of spacings, the high resolving power of which it is capable, the large aperture and the freedom from ghosts. It is only in the ultra-violet that the two first-mentioned instruments possibly still offer some

comparative advantages, owing to the difficulty of making good partially reflecting surfaces for this range of the spectrum.

The main developments in the technique of the single etalon since Fabry and Perot's early work have been the improvement of the reflecting surfaces, some studies of the practical performance of the instrument and the use of photoelectric recording.

We assume at first that the surfaces of the etalon plates are perfectly plane and parallel. The values R and T (for unequal films each of these values has to be taken as the geometrical mean for the two plates), together with the spacing D, then completely determine the properties of the instrument. The performance can be characterized by (*a*) the half-value width of the fringes, (*b*) the contrast factor and (*c*) the maximum brightness, in the same way as it was done for interference filters, and the same equations (3), (4) and (5) can be applied. In using equation (1), we now have to assume the wavelength to be constant, i.e. the light to be strictly monochromatic, while the variation of Δ is caused by the factor $\cos\phi$.

If the wavelengths of two equally strong spectral lines differ by an amount equal to the value $\Delta\lambda_{1/2}$ as defined on p. 72, equation (1) shows that the sum of their intensities is a function with two maxima separated by a depression of roughly 20%. If we define this condition as the limit of resolution, the resolving power is found to be $\lambda/\Delta\lambda_{1/2}=nN$, with the definitions given above. While the order n depends only on the spacing D, the factor $N=\pi\sqrt{R}/(1-R)$ is entirely a property of the reflecting surfaces. Two spectral lines of equal intensity appear resolved if they are separated by one-Nth of one order.

The value of R alone also determines the contrast factor, according to (5). For reflectivities above 0·8 as used in practice, it is approximately given by $I_{max}/I_{min}=0{\cdot}45N^2$.

This factor indicates how far the instrument is capable of detecting faint satellites. Its reciprocal is about equal to the ratio which the intensity of a satellite just detectable bears to the intensity of the main line, under the condition that the spacing has been so chosen that the satellite appears about mid-way between successive orders of the main line.

The brightness of the maxima depends on T as well as on R. Equation (4) shows that in the ideal case of an etalon with absorption-free films of equal R, the brightness in the maxima is always as great as that of the incident light. Under favourable conditions, silver films can approach this limit remarkably closely. As, however, the thickness of the metal film is increased, the value of R approaches a limit $R_\infty<1$ while T tends to zero, so that the brightness decreases and approaches zero. Therefore, the thickness of the metal film has to be chosen as a compromise between the conflicting requirements of resolving power and brightness. T usually serves as a convenient measure of the thickness, because it can easily be measured during the process of evaporation. Owing to the change of T with wavelength, considerably thicker films are required for blue than for red light.

The connection between T and R for silver films deposited under good 'standard' conditions is now well known and fairly reproducible. This often allows the measurement of R to be replaced by the very much simpler measurement of T. The decrease of R, for freshly deposited films, during the first few weeks is not quite negligible (of the order of 1%). Figure 6 shows N and I_{max}/I_0 as functions of T for 3 weeks old silver films (Kuhn and Wilson 1950). If, for

example, an intensity of the fringes of 0·35 of the intensity in the absence of the etalon is required for silver films several weeks after evaporation, Figure 6 shows that for red light one 90th of an order can be resolved, and for green light one 35th of an order. In making the silver films, one would have to continue the evaporation until a transmission factor of 0·02 for red light is reached (or 0·05 for green light, in the second case).

In actual practice, the imperfections of the etalon plates have a noticeable effect on the performance of the etalon, even for the best plates obtainable. The greater the reflectivity, the more highly accurate plates are required. Bright, Jackson and Kuhn (1949) measured the width of etalon fringes, choosing the conditions so that the light could be regarded as practically monochromatic. For not too high reflectivities, the values of N derived from measured widths of fringes agree well with the values derived from reflectivity measurements, but as N increases above 20, the former are found to lag more and more behind the latter. The smallest fraction of an order which could be resolved, for red light,

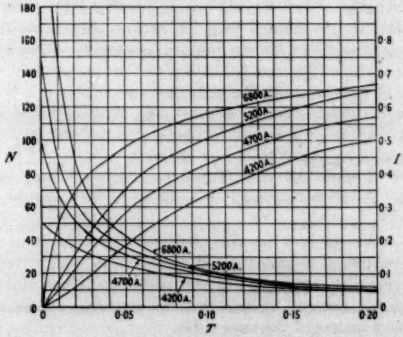


Figure 6. Resolving power and brightness of etalon fringes.

was about 1/50. It appears that values of $R>0{\cdot}94$ cannot be utilized with etalon plates of the quality at present available. For plates of large diameter, this limit is even lower. The surface profile of large etalons can sometimes be noticeably improved by the application of mechanical pressure by means of a subsidiary spring-loaded stud touching one plate in a suitably chosen point.

In order to get a qualitative picture of the influence of surface imperfections on the intensity distribution, one can assume that different parts of the etalon independently produce fringe patterns slightly shifted compared with one another. The result is the same as if, during an exposure, the photographic plate were slightly moved to and fro at right angles to the fringes. This leaves the flat portion of the intensity distribution curve near the minimum practically unaltered, while broadening and consequently lowering the peaks. Dufour and Picca (1945) expressed this assumption in a mathematical form and confirmed the results by some measurements. The assumptions are only valid if the variations across the surface are sufficiently gradual and the angles of incidence small enough.

If the light is not strictly monochromatic, a similar effect will be observed; the different wavelengths will produce curves slightly shifted in relation to one another, thus lowering the contrast ratio and the maximum intensity.

For a calculated contrast ratio of 200, the value of this ratio measured with sufficiently monochromatic light was only a little lower ($\simeq$150), while for a calculated value of 500 only about half of this value was found by experiment (Bradley and Kuhn 1951). It appears that, without excessive loss of intensity, contrasts much greater than 300 cannot be achieved with a single etalon.

When using calculated values of brightness, such as those given in Figure 6, or of contrast factors, one has to consider that the intensity actually observed will be less than the calculated value if either the spectral line width or the 'imperfection width' ($\simeq$one 50th of one order) is of similar magnitude to the calculated width (one Nth of one order). Thus, if the thickness of the silver film is made greater than that required for the necessary resolution, the loss of intensity will be found to be even more than that shown by Figure 6.

For aluminium films, N can probably be made about 15 in the near ultraviolet and rather less for 2500 A. (Crawford, Gray, Schawlow and Kelly 1949), but no actual measurements of resolution appear to have been made. Above 4000 A., silver is certainly preferable to aluminium.

Several authors have suggested increasing the reflectivity of the metal coatings in an etalon by covering them with one or two films of a dielectric, and also replacing the metal films entirely by multiple dielectric films, but no results obtained by either of these methods have yet been reported. Especially with the latter type of film, which is several wavelengths thick, it remains to be seen if the necessary flatness of the effective reflecting surface can be achieved. At present, dielectric films appear to be most promising for purposes where great brightness but not excessive resolving power is essential.

Dufour (1945) investigated the properties of an etalon having an opaque silver coating on one plate, with a 'window' allowing the light to enter. If only small apertures are required, he finds that the method offers advantages over an ordinary etalon. This method was first used by Barnes (1904).

Tolansky and Ranade (1949) suggest the use of reflection fringes for spectroscopic work with faint sources. But since the brightness of the ordinary transmission fringes is of the order of one quarter that of the incident light, with suitable choice of the film thickness, a factor of about 4 is the upper theoretical limit, set by the second law of thermodynamics, for gain of brightness by any method. A possible gain of 30 mentioned by these authors therefore seems excessive. Moreover, the drawback that the reflection fringes are dark fringes on a bright background will more than outweigh any possible small gain in intensity.

§ 9. THE DOUBLE ETALON

The method of using two etalons in series, developed by Houston (1927) and Gehrcke and Lau (1927, 1931) has often been described, but surprisingly few investigators have actually used it. Too low values of the reflectivity of silver which are often quoted have led to an overestimate of the loss of light and of the difficulties caused by unsuppressed fringes. Since also in other respects the treatments in the textbook literature are incomplete or even misleading, some comments based on practical experience with the instrument may be justified.

For monochromatic light, the theory of the double etalon is very simple. Each etalon acts as an 'angular filter', the transmission of which is given as a function of the angle of incidence $I(\phi)$ by equation (1). Two etalons in series, sufficiently well aligned, therefore have an intensity distribution (transmission function) which is the product of the functions $I(\phi)$ of the individual etalons. The complicated calculations given by Houston (1927) and Meissner (1942) are in no way more rigorous and have often led to the misconception that some interferometric action between the two etalons takes place. In fact, reflections between the two etalons lead to undesirable effects and have to be excluded (Bradley and Kuhn 1951). The treatments mentioned are so complicated only because they ignore the fact that several parallel coherent waves are mathematically identical with one single wave. The action of the first etalon is therefore exhaustively described by the intensity distribution function $I(\phi)$.

It is by no means necessary to make the ratio of the two spacings an exact integer. The condition for the choice of the two spacings D and D' is $D' = mD \pm d$, where m is a small integer, generally between 1 and 10, and d is an integral multiple of $\lambda/2$ and should be smaller than D/N, where N is the effective number of beams. The latter condition means that the scale factor of the two etalons should be in an integral ratio to an accuracy within $1/N$. It is generally easy to fulfil this latter condition.

If d is finite, the 'phase adjustment' (i.e. making d an integral multiple of $\lambda/2$) is different for different wavelengths. In white light, the spectrum in the focal plane of the spectrograph is broken up into bright and dark bands (Edser–Butler fringes); if the double etalon is correctly in phase for any particular line to be investigated, the maximum of a bright band appears in the position of the line. The double etalon then acts as an interference filter of thickness $|D' - mD|$. This effect can be useful for the suppression of strong, adjacent lines which are insufficiently separated by the dispersive action of the prism spectrograph.

The phase adjustment can be carried out in several ways. Houston's original method of tilting one etalon relative to the other is somewhat crude and can only be applied with small spacings when fringes far from the centre can be used. If the contact surfaces of the spacers are made sufficiently curved, variation of the spring pressure allows a very close phase adjustment which can be followed by a finer adjustment by means of tilt. Roller bearings, for example, which are very suitable for not too large spacings, allow this kind of adjustment. If perfect stability during long exposures is required, this method is not always satisfactory, and, especially for large spacings, the method of phase adjustment by change of air pressure (Jackson and Kuhn 1938) must be used. The inconvenience of using an airtight box with optically good windows is compensated by the gain in temperature stability and protection of the silver surfaces.

The fact that the 'suppressed fringes' are not completely absent must, of course, be borne in mind, but since their positions are accurately known, they can hardly lead to errors: clearly, one would not look for a faint spectral line in a position coinciding with a suppressed fringe of a much stronger line. The intensity of a suppressed fringe can be calculated from the reflectivity of the films by means of equation (1) or can be measured by means of a suitable light source, such as a pure isotope mercury lamp.

The incomplete suppressor action of the smaller etalon limits the ratio of spacings m which can be used. Generally speaking, a ratio up to 4 can be used

throughout most of the visible spectrum, but in the red, a ratio of 10 has been used with success (Jackson and Kuhn 1939).

The maximum transmission factor of a double etalon is the product of the factors for the individual etalons. But the loss of brightness caused by the addition of the second etalon can often be compensated by the choice of slightly thinner silver films, and the view that double etalons can only be used for strong lines is incorrect. In a recent investigation of a very weak component of the hydrogen α line, emitted by a discharge carrying 6 ma. only (Series 1950), replacement of a single by a double etalon led to a considerable gain in resolution with an increase of exposure time of only about 50%.

The main purpose of the double etalon with different spacings is the increase of the range between orders, allowing complicated patterns to be investigated. A different effect is achieved if two etalons of equal spacings are used in series. Theoretically the contrast factor of this combination should be the product of the individual factors and therefore extremely large. Bradley and Kuhn (1948, 1951) investigated the possibilities of this method and used it as a means of detecting faint satellites of helium lines due to the isotope ^{3}He, present in a concentration of 1/700.

In attempting to produce great contrast ratios, it is, of course, essential to exclude false reflections very rigorously. The reflections from the unsilvered surfaces can easily be avoided by suitable orientation of the wedge angles of the plates. The reflections between the silvered faces of the two different etalons present a much more real difficulty. This can be overcome by making the distance between the two etalons sufficiently large compared with their aperture, and by avoiding the use of fringes too close to the centre. Figure 7 shows how, in this way, the back reflection from the first etalon can be made to miss the second etalon.

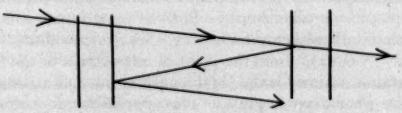


Figure 7. Prevention of back reflection in the double etalon.

For green light, a contrast ratio of about 2,000 was actually measured. Though this value was much less than the calculated value, it is about 10 times as great as the contrast in a single etalon.

§ 10. PHOTOELECTRIC RECORDING

In many branches of spectroscopy, the development of photoelectric cells and especially photomultipliers has raised the question how far direct photoelectric recording can replace the photographic plate. The method offers special advantages for intensity measurements, but it depends largely on constancy of the light source. Jacquinot and Dufour (1949) describe a photoelectric recording technique which utilizes the light contained in a number of orders of the etalon fringe system. The fringes are projected on a screen which has a system of circular, transparent zones, made by photographing etalon fringes of a monochromatic source. The etalon is enclosed in an airtight chamber and slow

continuous variation of the air pressure causes the fringes to pass over the system of circular zone diaphragms. The light passing through the latter is focused on the photomultiplier. The galvanometer deflection is recorded on a drum whose rotation is synchronized with the change of air pressure.

In order to avoid using different zone diaphragms for different wavelengths and different spacers, Jacquinot and Dufour use a compound objective of variable focal length for the formation of the fringes.

§ 11. SOLAR INTERFEROMETRY

Fabry and Perot (1902) had already applied their interferometer to problems of absorption spectra which are typical of astrophysics. In their original arrangement, the prism spectrograph merely serves as a monochromator, and the light emerging from the exit slit is analysed by the interferometer. This procedure is wasteful of light and it allows only one line to be photographed in any one exposure. These drawbacks were avoided in an arrangement also due to Fabry and Perot (1902, 1905) and developed mainly by Shane (1932, 1941), in which the light passing through the etalon is focused on the entrance slit of the spectrograph. The resulting heterochromatic fringes appearing in the focal plane of the spectrograph have the shape of parabolae. Any imperfections of the spectrograph, however, affect the definition of the interference fringes; further, the aperture, generally small, of the high dispersion spectrograph is apt to limit the number of interfering beams in the etalon and thus to cause loss of resolution.

Treanor (1949) has developed a promising new method which appears to avoid the drawbacks of the two preceding methods. The light emerging from the focal plane of the spectrograph is collimated (without a slit), passes through the etalon and is then focused on the photographic plate. The heterochromatic fringes thus produced differ in shape from those in Shane's method, since the etalon is no longer ' centred ' except for one wavelength. Both by experiment and by calculation, Treanor shows that the heterochromatic fringes are circles whose centre is displaced by approximately $f^2(d\lambda/dx)/\lambda$ from the point of intersection of the focal plane with the axis of the etalon, where f is the focal length of the collimating lens and x the coordinate on the photographic plate. The approximation holds as long as the dispersion can be regarded as constant. If, either as the result of an increased slit width or of the limited resolution of the spectrograph, light of a given wavelength is spread over a finite angle when entering the etalon, the latter will cause this light to produce short arcs of circles forming part of the ordinary monochromatic fringe pattern, with its centre on the axis of the etalon. Narrow absorption lines will thus appear as arcs of intersection of the various orders of the monochromatic ring pattern with the heterochromatic displaced ring pattern.

This method of ' circular channels ' can be applied, in a manner similar to that of parabolic channels, to the accurate determination of wavelengths. The method of circular fringes lends itself especially to the study of line profiles. In either method the spacing of the etalon is limited by the resolving power of the spectrograph if overlapping effects are to be made negligible: the slit pattern of the spectrograph must be well within the spectral range (distance between adjacent orders) of the etalon, the existence of extended wings in a slit pattern making this restriction especially serious. In spite of these limitations, Treanor finds, by experiment and by a detailed analysis of these factors, that with silver coatings of only very moderate

reflectivity, a very considerable gain of resolving power over that of the spectrograph alone can be obtained.

Treanor points out that the newly developed methods of altering the diffraction pattern of the spectrograph, known as apodization (Boughon, Dossier and Jacquinot 1946, 1950) might help to increase the available resolving power. Another possibility would be the application of the double etalon.

III. STANDARDS AND TESTING METHODS

§ 12. LIGHT SOURCES AND WAVELENGTH STANDARDS

Interferometric methods are the most accurate means, not only of comparing the wavelengths of different spectral lines, but also of comparing mechanical gauges and standards of length with one another and with wavelengths of spectral lines. The latter use especially requires a source emitting light which is as highly monochromatic as possible and of constant wavelength. Most present legal standards of length are ' line standards ' defined by the distance of two fine, ruled lines from one another. The increasing use of interferometric methods has tended to replace line standards and line gauges by ' end-gauges ' in which the length is defined by the distance of two nearly parallel, highly polished metal surfaces. The ultimate aim is the replacement of the present legal standards by one single wavelength standard.

It is generally recognized that Michelson could hardly, at his time, have chosen a better wavelength standard than the cadmium red line, but various possibilities of better standards now exist. The width of the cadmium red line as emitted by an ordinary discharge tube is mainly determined by the Doppler effect. The use of a heavier atom as emitter and cooling of the discharge appear to be the obvious ways of making a more highly monochromatic source. The cooling could be effected easily with a gaseous emitter or by the use of a hollow cathode tube. But the success of such methods would be severely limited by the hyperfine structure which most spectral lines possess.

This structure arises from two causes: (1) Isotopes of odd atomic weight have a nuclear spin causing splitting of most lines, and practically all single-isotope elements have odd atomic weights. (2) Elements consisting of several isotopes show a structure in most spectral lines which is due to the effect of ' isotope shift '. In the cadmium red line, it is the latter effect which sets a limit to the decrease in width resulting from cooling the source. Krypton lines, which are otherwise very suitable as standards, suffer from the same drawback.

Pure isotopes of many elements are now becoming available in sufficient quantities for use in discharge tubes, and great improvements of wavelength standards have thus become possible. The mercury isotope 198 can be made in a pure form by neutron bombardment of gold. Wiens and Alvarez (1940) and Wiens (1946), exposing gold to neutrons from a cyclotron, were the first to obtain quantities sufficient for a discharge tube to be run for a short time. Milligram quantities of the isotope can now be made in uranium piles. Meggers (1948) discussed the possibilities of using ^{198}Hg sources as standards. An important technical difficulty is the loss of the metal in discharge tubes. It is caused by sputtering of the electrodes and, in electrodeless discharges, by other processes driving the mercury into the glass walls. The General Electric Company have

succeeded in developing discharge tubes with internal electrodes which have a life of several hundred hours with a charge of only one-tenth of one milligram of ^{198}Hg. Even without the use of water cooling, the green line emitted by these discharge tubes is not only appreciably narrower than the cadmium red line but is also considerably brighter.

For purposes of high resolution spectroscopy, the mercury isotope lamp is of greatest value for adjustment of etalons with large spacings and for the determination of the true instrumental intensity distribution, contrast factors and for similar purposes.

Accurate comparisons of the wavelengths of the lines of ^{198}Hg with the wavelength of the cadmium red line have been carried out by the National Bureau of Standards and by the National Physical Laboratory. The results are given in the Table.

Wavelengths of ^{198}Hg lines in dry air at 15° C. and 760 mm. Hg pressure, referred to a value of the Cd red line of 6438·4696 A.

Line	N.B.S.	N.P.L.
Yellow 1	5790·6628	5790·6628
Yellow 2	5769·5983	5769·5985
Green	5460·7532	5460·7531
Violet	4358·3373	4358·3376

Mercury isotopes in an almost pure form can also be made by electromagnetic separation from ordinary mercury. McNally, Griffin and Burkhart (1949) describe the spectroscopic use of the isotope 202; having an even mass number, it is also free from hyperfine structure and offers the same advantages as the isotope 198.

The new methods developed by Koch (1948) for the separation of isotopes of the rare gases might also allow the lines of the pure isotopes of krypton to be used as wavelength standards.

It is expected (Barrell 1948) that standards of length established by means of pure isotope light sources should be reproducible to an accuracy within 10^{-8}, compared with an accuracy within 10^{-7} as the limit which might be achieved with ordinary line gauges. If still greater accuracy is required, the Doppler width would have to be reduced by replacing the emitting gas or vapour by an atomic beam. Such light sources have been used for hyperfine structure research, but their low intensity would present a difficulty in their use as wavelength standards.

In the comparison of interferometrically measured standards, it is necessary to apply corrections for differences in atmospheric pressure and temperature, and these involve the refractive index of air. New measurements by Pérard (1934), Kösters and Lampe (1934) and Barrell and Sears (1939) have given values of this quantity differing noticeably from older values and probably considerably more accurate than these.

As secondary wavelength standards, krypton lines are frequently used. Littlefield (1946, 1950) applied a reflection echelon to a determination of krypton wavelengths relative to the wavelength of the cadmium red line. Though the reflection echelon has, quite apart from the high cost, severe drawbacks such as the impossibility of varying the spacing, the low dispersion and the unfavourable intensity distribution, it appears to have some advantages for this particular purpose; no phase correction for the reflection is required and the instrument can be used over a wide range of wavelengths.

§ 13. GAUGE TESTING METHODS

The most common type of interferometer used for gauge measurements appears to be the Kösters interferometer (Kösters 1926, Weber 1928) which differs from the Michelson or Twyman–Green interferometer in design but not in principle. The length of gauge for which this type of instrument can be used depends only on the spectral purity of the light. The extreme limit of the optical path $2D$ for which fringes are observable with a light source of half-value width $\Delta\lambda$ is $2D = \lambda^2/\Delta\lambda$.

With ordinary krypton lamps, differences up to 12·5 cm. were measured in a determination of displacements of the piston in an adjustable cavity resonator (Barrell and Puttock 1950), but isotope lamps should allow the direct measurement of much greater distances.

For smaller gauges, Fizeau's method can be used, with the polished end-face of the gauge acting as reflector and a lightly aluminized glass plate as beam splitter. Barrell (1948) describes an interferometer of this kind in use at the National Physical Laboratory and the details of the procedure.

An interesting modification of the Michelson interferometer was developed by Peck (1948), and independently at the Zeiss Works (Räntsch 1949). The movable mirror is replaced by two plane mirrors, at right angles to one another, meeting in a vertical edge, and the light wave suffers two reflections at angles of incidence of 45°. Rotation of this reflector about the edge as axis leaves the fringes unaffected. If it is mounted on one end of an arm, the other end of which is pivoted, angular displacements can be measured interferometrically. Sittner and Peck (1949) used this method in a recording infra-red spectrometer.

§ 14. TESTING OF WAVE FRONTS

The main purpose of the Twyman–Green interferometer is not merely to test the geometrical shape of a surface, but to test wave fronts emerging from optical systems when used with a point source. Any deviation from the spherical (or plane) shape which would cause aberration in the image can be tested interferometrically by this instrument (see Williams 1941). An example of some recent work is the study of aberrations of telescopic systems by Coleman *et al.* (1947).

Interferometric tests of this kind require a practically perfect reference wave front, and therefore a practically perfect mirror or lens of the same size as that under test. In the testing of large astronomical mirrors, this requirement is difficult to fulfil. With the object of avoiding the reference wave front, Bates (1947) designed the 'wave front shearing interferometer' in which a sheared image of the wave front itself is used as reference wave front. By an arrangement of two plane reflectors M_1 and M_2 and two beam dividers D_1, D_2 (Figure 8) very similar to that in the Mach–Zehnder interferometer, the single incident wave front is divided into two coherent wave fronts of exactly equal shape, which can be made to coincide with one another. With a point source, the mirror or lens to be tested produces a converging wave front, forming a point image near D_2. The eye placed at E will receive waves appearing to come from the two wave fronts F_1 and F_2. A slight tilt of D_2 will cause a shear of F_1 and F_2 relative to one another, combined with a tilt, if the image point is outside D_2. Also the image points P_2, P_1 of the two waves will no longer coincide. The eye placed at E and looking at a distant plane such as that of the reflected image of the mirror under test, will see interference fringes;

if the incident wave front is perfectly spherical, they will be approximately straight and at right angles to the plane of the paper; any deviation from sphericity will appear as distortion of the fringes.

If the aberrations are of axial symmetry, it is possible to derive the shape of the wave front from the interference patterns caused by different amounts of shear of the two wave fronts F_1 and F_2. A modification of the instrument allowing rotation

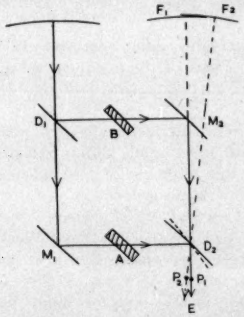


Figure 8. Bates' wave front shearing interferometer.

of the wave front as well as shear relative to one another has even extended the possibilities of the method.

The two glass plates A and B are required to compensate the effect of the glass plates carrying the partially reflecting surfaces D_1 and D_2. Since the fringes are of low order, white light can be used, and they can be seen with a slit source or with a star serving as point source.

This interferometer has been used for the detection of zonal errors, in the figuring stage, of the 74 in. mirror now installed at the Radcliffe Observatory at Pretoria.

§ 15. TESTING OF INTERFEROMETER PLATES

In the methods described in the preceding paragraphs, two beams only are made to interfere, causing a $\cos^2$ intensity distribution in the fringes. The accuracy of such tests is therefore difficult to increase beyond a limit of about 1/30 of one wavelength. While this accuracy is ample for ordinary, image-forming systems, it is insufficient for the surface of an etalon plate. The surface errors in a good etalon do not exceed 1/50 of one wavelength over distances of several centimetres, but a considerably higher accuracy would be highly desirable. It is clear that for plates which are used for multiple beam interference, a satisfactory test can only be carried out by the use of multiple interference fringes. In the manufacture of etalon plates, multiple beam testing requires frequent re-coating of the plates. It is perhaps owing to this inconvenience that the method appears to be so little used.

Two different techniques can be used for multiple beam testing of etalon plates. Duffieux (1932) and Rasmussen (1945) place the two plates opposite and fairly close to one another, so that they form an air wedge of small angle. Fringes of

equal thickness are then observed or photographed in collimated monochromatic light. The arrangement differs from that of Fizeau only in so far as transmitted light is observed instead of reflected light. The fringes appear as extremely fine lines, and very slight variations in the spacings can be detected by measurement of the photographs. Rasmussen claims to be able to detect errors of 1/600 of one wavelength.

The other, less laborious, method is that which is in common use for adjustment of etalons. It can be made more quantitative, if, instead of the naked eye, a low power telescope is used and the diameters of the fringes are measured against an eyepiece scale. The entrance pupil must, of course, be kept sufficiently small. Even this simple method is far more sensitive than a Fizeau test with un-silvered surfaces, and sufficiently sensitive for the detection and location of errors in the best commercial etalon plates.

IV. THE STUDY OF SURFACE STRUCTURES

§ 16. THE PRINCIPLE OF THE METHOD

The interferometric examination of surfaces has long been a well-established practice, but the improvement of this method by means of multiple reflections, first proposed by Fabry, has only found occasional application (see, for example, § 15) until recently. Tolansky appears to have been the first to recognize the potentialities of the method when used in conjunction with the modern technique of vacuum evaporation. He and his co-workers developed it in several forms and applied it extensively to the study of surface structures of crystals (Tolansky *et al.* 1943–1949).

Fringes of equal thickness (Fizeau fringes) located in the surface itself are always used. A light source of restricted size, placed in the focal plane of a lens, produces a collimated beam. The fringes are formed by multiple reflection, at nearly normal incidence, between two silver-coated surfaces, one being the surface to be studied, the other the plane surface of a glass plate placed close to the latter. Owing to the occurrence of multiple reflections, the fringes are very narrow and therefore clearly defined, and follow the contour lines of equal distances between the reference flat and the surface to be studied.

Tolansky and his co-workers use four methods of observation: (i) fringes in transmitted monochromatic light, (ii) fringes in reflected monochromatic light, (iii) heterochromatic fringes in transmitted and (iv) heterochromatic fringes in reflected light.

Methods (iii) and (iv) differ from (i) and (ii) less than might appear at first sight, since the actual analysis of the contour is always carried out for monochromatic light produced, in methods (iii) and (iv), by the action of a spectrograph on white light. The remarks in this section therefore apply equally to any of these methods.

In the ideal case, the phase difference Δ between two successive rays depends only on the distance D as the sole variable, and all that was said in Part I about the intensity distribution $I(\Delta)$ applies directly to these fringes. But how far the ideal case is realized in practice depends on the following factors: the wedge angle ϑ, the absolute value of the spacing D and the degree of collimation of the light as determined by the ratio γ of the diameter of the source to the focal length of the collimating lens. As in ordinary Fizeau fringes, γ must be the smaller the larger D, if blurring of the fringes is to be avoided, but the smaller width of the multiple

beam fringes makes this condition more critical. If the light is not sufficiently collimated, the fringes are asymmetrical and appear shaded off in the direction of increasing D, with a sharp edge towards decreasing D. This is due to the factor $\cos\phi$ contained in the term Δ in equation (1) which has the value 1 as an upper limit. This asymmetry can often be useful for distinguishing between upward and downward slope of the surface.

Another cause of fringe broadening depends exclusively on the properties of the wedge: after a beam has suffered a large number of reflections, it has not only drifted some distance away from the point of incidence but has also changed its direction. This makes the interference condition different from that of the ideal case and causes loss in definition. This effect sets an upper limit to the value of D which can be used for any given wedge angle. Brossell (1947) has studied the influence of the degree of collimation of the light and especially of this latter 'wedge effect' on the appearance of the fringes. If $X=2\vartheta/\lambda$ is the number of fringes per cm., the condition that the nth reflected beam has suffered a phase lag of π as a result of the wedge effect, is approximately

$$D=3/2n^3\lambda X^2. \qquad \ldots\ldots(6)$$

If the effective number of beams, N (see §5) of the order of magnitude 50, is substituted for n, this formula gives an estimate of the upper limit of D; it is the smaller the more closely the fringes are spaced.

§17. FRINGES IN MONOCHROMATIC LIGHT

For transparent materials, it is most convenient to observe fringes in transmitted light. Both the surface to be tested and that of the reference flat are coated with silver films of reflectivity about 0·8 to 0·9. Collimated light is made to enter the crystal, and fringes are observed through the reference flat by means of a magnifying glass or low power microscope, either visually or photographically. Since the fringes are of low order, the source need not be highly monochromatic; a high pressure mercury lamp, in conjunction with a colour filter, is a convenient, bright source.

The microscopic observation raises the question, how far the resolving power in depth is connected with the lateral resolving power which can be achieved simultaneously. Since the spacing of the fringes, for a wedge angle ϑ, is $\lambda/2\vartheta$, the use of N effective beams should make the half-value width of a fringe

$$\delta' x=\lambda/2N\vartheta=\lambda/(\alpha/2), \qquad \ldots\ldots(7)$$

where α is equal to the angle which the Nth beam forms with the directly transmitted beam; it is, in fact, the angular spread caused by multiple reflection. A microscope will be able to resolve this width $\delta' x$ only if it has a sufficient aperture to receive the Nth beam. Equation (7) agrees in order of magnitude with the Abbe formula for the resolving power of a microscope of numerical aperture $\alpha/2$. But this estimate, which Tolansky gives in his book, represents a limit which will be hard to approach in practice. In the numerical example, in this book, of fringes with a spacing of 0·1 mm., the critical value of D is 1/5 wavelength, which could only be realized under most exceptional conditions. With values of D equal to or greater than λ, as occur in practice, the lateral displacement of the rays resulting from multiple reflections becomes very noticeable. The displacement of the Nth ray compared with the first is

$$\delta'' x=2N^2D\vartheta=ND\alpha/2. \qquad \ldots\ldots(8)$$

If α is made smaller than the aperture of the microscope objective the fringes are, according to (7), wider than the resolving limit of the microscope. If, however, α is as large as the latter, condition (8) will, with large values of N, often make $\delta''x$ greater than $\delta'x$. A rough estimate of the maximum aperture can be made by equating (7) and (8), with the result $\alpha = 2(\lambda/ND)^{1/2}$. Assuming that D cannot be made smaller than λ, this condition becomes $\alpha \leqslant 2/\sqrt{N}$.

This estimate, though only aiming at the order of magnitude, shows that the gain of resolution in depth by the use of multiple reflections is necessarily connected with a loss in lateral definition. This loss becomes greater as D increases.

If the substance whose surface is to be studied is not sufficiently transparent, the Fizeau fringes have to be observed in the reflected beam. The experimental arrangement is very similar, but a partially reflecting glass plate, set at 45° to the beam reflects the collimated light on to the reference surface at nearly normal incidence, while allowing direct observation of the reflection fringes through the inclined plate.

As mentioned above, the reflected light shows dark fringes on a bright background, but otherwise the use and interpretation of the fringe patterns is the same as for transmitted light. The technique is, however, more difficult in several respects: if the reflectivities of the two surfaces are not well matched, the fringes lack contrast; they can also be asymmetrical (see § 5) for reasons other than those discussed in § 16; also, microscopic observation is made less convenient by the need of a reflecting plate between microscope and object.

Timms (1945), studying the surface structure of polished metals by means of multiple beam fringes in reflected light, used a cylindrical reference surface.

At oblique incidence of the light, the Fizeau fringes produced between two semi-silvered glass surfaces appear double, owing to the difference of the phase shifts for the two planes of vibration. From his measurements of this effect, Tolansky (1944 a, b) derived the difference of these phase shifts and compared it with the value calculated by MacLaurin (1906) from the optical constants of silver.

§ 18. HETEROCHROMATIC FRINGES

If, in the experiment of Edser and Butler, the air plate in front of the slit is made wedge shaped (with the edge of the wedge horizontal and the slit vertical), the fringes produced in the spectrograph by white light are inclined. The intensity distribution along any vertical line in the focal plane is that of interferences of equal thickness produced in monochromatic light. It is determined by the variation of the gap between the two plates, along the line of the slit. The intensity distribution along any other vertical line is the result of the same ' contour ' observed in light of a different wavelength.

This is the arrangement used by Tolansky in producing his ' fringes of equal chromatic order ', though the air film formed by the surface of the specimen and the reference plate is not placed directly in front of the slit but projected on the slit by means of a lens, either by the transmitted or the reflected light. This method thus gives the contour along one line only, and not that for an area as does the technique using monochromatic light. The direction of the slope of the heterochromatic fringes shows immediately the direction of increasing spacing. This is apparently the main advantage of the method. In some cases, Tolansky finds it preferable to the monochromatic method, in spite of its limitation to a line contour.

§ 19. TECHNICAL DETAILS AND APPLICATIONS

The spacing of the fringes, in any of the methods described, can be adjusted by the tilt of the reference plate and has to be adapted to the particular purpose. Widely spaced fringes are very sensitive to differences in height, but the large areas between the fringes are missed by the analysis. In order to combine the advantages of small and large fringe spacings, Tolansky superimposes several fringe patterns of the same surface area, by taking successive photographs with different tilts of the reference plate.

With the use of very widely spaced fringes ('high dispersion fringes' in Tolansky's nomenclature) the method reaches its highest sensitivity and should be able to detect level differences of only a few Ångström units. But even the best etalon surfaces are only flat to about 50 A. (probably rather better over small areas). The need of a reference flat will therefore make it difficult to utilize the high potential sensitivity of the method. Conditions are specially favourable in the measurement and detection of abrupt changes in height. Tolansky was able to find that cleavage steps in mica appear as multiples of a unit of 20 A. This value does not agree with the results of Marcelin (1931) who found these steps to be multiples of 7 A. The results of the multiple beam technique appear likely to be more accurate than those of the double beam technique, even in the refined form which Marcelin uses.

If an area forms part of a concave surface or is otherwise inaccessible to direct interferometric study, replica techniques have to be applied. Faust and Tolansky (1947) found that methylmethacrylate gave satisfactory results, reproducing details of the surface to an accuracy within 10 A. in depth.

The results of applying the methods of multiple beam interference to the study of crystal surfaces are outside the scope of this report. The method has especially been applied to mica, selenite, quartz, calcite and diamond and has greatly contributed to the knowledge of the surface structure of crystals.

In the investigation of metal surfaces, the scope of the method is similar to that of the phase contrast microscope, but it surpasses the latter in resolving power 'in depth'.

An interesting more recent application (Tolansky and Bardsley 1948) is the study of modes of vibration of crystals. The vibrations cause the fringes to be blurred, except in the vicinity of the nodes which can thus be located.

V. INTERFERENCE REFRACTOMETERS AND INTERFERENCE MICROSCOPES

§ 20. REFRACTOMETRY OF HOMOGENEOUS MEDIA

Interferometric methods are generally used for measuring refractive indices of gases and vapours and for measuring very small changes of refractive index in liquids, such as may be caused by small changes in composition. The two most important instruments, the Jamin and the Rayleigh refractometers, are described in most textbooks, but their relative advantages and drawbacks require a brief discussion.

A great advantage of Jamin's instrument is the brightness of the fringes and their convenient size, allowing them to be viewed through a telescope of low power. The two thick, optically homogeneous and closely matched glass blocks, however, make the instrument expensive. Moreover, these glass blocks are apt to cause

drift of the fringes, unless the temperature is kept constant over long periods. In the Rayleigh refractometer, difficulties arise from the wide spacing of the two beams which is required of any refractometer. This makes necessary the use of a very narrow entrance slit which leads to loss of light and causes the spacing of the interference fringes to be microscopically small. The excessive magnification of the eyepiece thus made necessary causes further loss of light. The small scale of the fringes makes it impracticable to use a cross hair as reference mark, because minute relative displacements of fringes and cross wire would cause noticeable errors. The complicated device for producing a reference fringe system, often described as a special advantage of the instrument, is merely a necessity, arising from the extreme smallness of the fringes.

Whereas the disadvantages of the Rayleigh refractometer are inherent in the method of diffraction which does not lend itself readily to producing a wide separation of beams, the drawbacks of the Jamin instrument are of a more technical nature. They can be avoided if the glass blocks are replaced by ' air plates ' formed by metal-coated glass plates of normal thickness. Kuhn and Wheatley (1945) have described an instrument in which commercial ball bearings are used as spacers in the ' air plates '. Experience gained with instruments of this type has shown them to be very suitable for accurate measurements. The glass plates need only be of moderate optical quality.

Another modification of the Jamin refractometer has been described by Lotmar (1949).

Arnulf and Cagnet (1950) used a Fabry–Perot interferometer as a refractometer by allowing the to light pass through the two halves of the aperture of the etalon in succession.

In interference refractometry, it is customary to use white light fringes and to compensate the differences of the optical paths by means of a Jamin compensator. The difference between the dispersion constants of the substance to be investigated and the glass of the compensator limits the range of this method. Craven (1945) discussed these limitations, showing that the direct measurement with white light gives the ' group refractivity ' rather than the ordinary refractive index.

§ 21. APPLICATIONS TO AERODYNAMICS

Optical methods offer obvious advantages in the study of local variations of density in liquids and gases such as occur as the result of high speed flow or of temperature variations. In supersonic flow especially, the variations of density are appreciable and allow fine details of the flow to be studied by optical means. Most of the methods are the same as those already used by Mach, but interest in them has increased lately, and some of the recent work is briefly discussed in this section. Three optical methods are in use for aerodynamic studies : the 'shadowgraph' method, the ' schlieren ' (striae) method and interferometry. In the first, the effects observed depend essentially on the second derivative of the density, and it therefore lends itself to the observation of abrupt variations. The schlieren method depends mainly on the first derivative of the density. The interference method is much more quantitative since it allows the density itself to be measured.

For observation and measurement of local variations of density, it is necessary to produce interference fringes located in or near the plane of the object under study. In Jamin's interferometer the fringes appear at infinity, but if the surfaces of one of

the blocks are not quite parallel to one another the fringes are produced at a finite distance. In Mach and Zehnder's modification of this instrument, the four reflecting planes can be adjusted independently, and the fringes can be produced in any desired position, thus allowing superposition of the interference pattern and the object plane. If the refractive index is uniform, the fringes can be made to appear as practically straight lines.

If a stream of gas is made to pass through one of the two interfering beams, at right angles to it, the fringes are displaced and distorted. The displacement at each point of the pattern is proportional to the change in optical path, integrated along the beam. If the flow has axial symmetry, an analysis of the interference pattern allows the change of density at each point of the flow, and thus the local pressure, to be calculated.

Ladenburg, Winckler and Van Voorhis (1948, 1949) describe interferometric studies of supersonic air flows and Ladenburg and Bershader (1949) investigate laminar and turbulent boundary layers. A thorough treatment of the formation of fringes in the Mach–Zehnder interferometer was given by Kinder (1946), and the method of applying this interferometer to aerodynamic studies was described in detail by Winckler (1948) who gives a full list of references.

The size of the available interferometer plates limits the area over which flow phenomena can be studied. Johnson and Scholes (1948) describe an interesting modification of the Mach–Zehnder instrument (Figure 9) in which only one

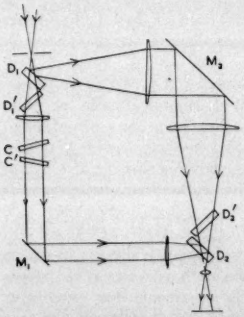


Figure 9. Johnson and Scholes' interferometer.

reflector M_2 and two lenses are required to be as large as the area to be investigated. The light is collimated after having passed through the first beam splitter D_1 and is made convergent before having passed through the second beam splitter D_2. The aberrations caused by the passage of non-parallel light through the glass plates D_1 and D_2 are the same in both beams and therefore do not affect the interference fringes. Only lateral chromatic aberrations have to be compensated by two glass plates D_1' and D_2', and the difference in optical thickness of the two lenses is compensated by C and C'.

In the schlieren method, interference fringes can also be observed under certain conditions and can make a more quantitative evaluation of the results possible (Gayhart and Prescott 1949).

§ 22. INTERFERENCE MICROSCOPES*

Any interferometer in which the fringes are localized in the space in which the beams are divided, allows local variations of refractive index to be observed, if necessary on a microscopic scale. Sirks (1893) and Pringsheim (1898) applied the term ' interference microscope ' to a Mach interferometer adapted for measurement of refractive index in very small areas, but ' micro-refractometer ' would have been a more appropriate description.

An ordinary microscope cannot reveal perfectly transparent objects differing from their surroundings in refractive index only: such objects change the phase of the light, but hardly the amplitude. Many biological objects and small differences of level in reflecting surfaces, such as metals, act in this way. Zernike's phase contrast microscope has become a powerful and comparatively simple means of observing objects of this kind. In this method, however, the use of diffracted light for producing interference with the image-forming light causes undesirable edge effects and makes quantitative work difficult.

These difficulties do not arise in interference microscopes. An instrument based on multiple beam interference was described by Frederikse (1933) and

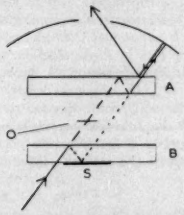


Figure 10. Dyson's interference microscope.

further developed by Merton (1947 a, b) and by Ambrose (1948); the method is in principle the same as that developed by Tolansky for surface topography and described in Part IV. Fizeau fringes are produced by multiple beam interference between two partially silvered glass surfaces, and the object is placed between them. If the wedge angle is made almost zero, local variations of refractive index are revealed as different shades in illumination by monochromatic light, or as colour differences if unfiltered light from a mercury discharge lamp is used.

Microscopic observation of multiple beam interferences in collimated light was discussed in §17, and it was seen that the use of multiple reflections restricts the aperture of the emergent beam which can be utilized. By similar arguments, it can be shown that the aperture of the incident beam is severely restricted. Merton mentions attempts which are being made to overcome these difficulties.

Systems based on double beam interference allow much larger apertures to be used. A relatively simple interference microscope has been developed by Dyson (1949 b) and is shown diagrammatically in Figure 10. Of two glass plates of equal thickness, one, A, is coated with semi-transparent silver films on both sides, and the other, B, on the upper side only, while the lower side is heavily silvered over a small circular area S. Every beam directed towards the object O is split into two

* The author is indebted to Dr. Barer for his advice on this subject.

beams (shown as broken and dotted lines) only one of which passes through O. Both beams have paths of equal length, as in the arrangement for producing Brewster fringes, and enter the objective as one resultant beam. The objective is of a special type (Dyson 1949 a) which allows a large working distance. The plate B is made slightly wedge-shaped, and lateral movement allows the phase difference to be adjusted in order to suit any particular part of the object. When white light is used, very small phase differences appear as colour contrasts.

Philpot (1948) has developed several new interference microscopes. In one of them, the interferometric part is designed as an attachment to be used with any microscope in place of the ordinary eyepiece.

Bates (1949) has suggested an interference microscope which is based on the same principle as the wave front shearing interferometer described in § 14.

REFERENCES

A. Books and Reviews

Barrell, H., *Light Waves as Standards of Length*, 1948, *Research*, **1**, 533.

Barrell, H., *New Developments in Interferometry*, 1949, *Nature, Lond.*, **164**, 599.

Born, M., 1933, *Optik* (Berlin : J. Springer).

Fabry, Ch., 1923, *Les applications des interferences lumineuses* (Paris : Édit. Revue d'Optique).

Jacquinot, P., and Rouard, P., *International Colloquium on Solid, Thin Films*, 1949, *J. Opt. Soc. Amer.*, **39**, 1048; see also *J. Phys. Radium*, 1950, **11**, 305–480*.

Meissner, K. W., *Spectroscopic Interferometry*, I, 1941, *J. Opt. Soc. Amer.*, **31**, 405; II, 1942, *Ibid.*, **32**, 185.

Tolansky, S., 1947, *High Resolution Spectroscopy* (London : Methuen).

Tolansky, S., 1948, *Multiple Beam Interferometry* (Oxford : Clarendon Press).

Williams, W. E., 1941, *Applications of Interferometry* (London : Methuen).

Wood, R. W., 1936, *Physical Optics* (New York : Macmillan).

B. Papers

Abeles, F., 1948, *Ann. Phys., Paris*, **3**, 504; 1949.

Ambrose, E. J., 1948, *J. Sci. Instrum.*, **25**, 134.

Arnulf, A., and Cagnet, M., 1950, *J. Phys. Radium*, **11**, 35 S.

Banning, M., 1947, *J. Opt. Soc. Amer.*, **37**, 792.

Barnes, J., 1904, *Astrophys. J.*, **19**, 190.

Barns, R. B., and Czerny, M., 1931, *Phys. Rev.*, **38**, 338.

Barrell, H., and Puttock, M. J., 1950, *Brit. J. Appl. Phys.*, **1**, 87.

Barrell, H., and Sears, J. E., 1939, *Phil. Trans. Roy. Soc.* A, **238**, 1.

Bates, W. J., 1947, *Proc. Phys. Soc.*, **59**, 940; 1949, Meeting of Brit. Assoc.

Billings, B. H., 1947, *J. Opt. Soc. Amer.*, **37**, 738.

Billings, B. H., and Pittman, M. A., 1949, *J. Opt. Soc. Amer.*, **39**, 978.

Blodgett, K. B., 1939, *Phys. Rev.*, **55**, 391; 1940, *Ibid.*, **57**, 921.

Boughon, P., Dossier, B., and Jacquinot, P., 1946, *C. R. Acad. Sci., Paris*, **223**, 661; 1950, *J. Rech. Cent. Nat. Rech. Sci.* (No. 12) 123.

Bradley, L. C., and Kuhn, H., 1948, *Nature, Lond.*, **162**, 412; 1951, to be published shortly.

Bright, R. J., Jackson, D. A., and Kuhn, H., 1949, *Proc. Phys. Soc.* A, **62**, 225.

Brossell, J., 1947, *Proc. Phys. Soc.*, **59**, 224.

Caballero, D. L., 1947, *J. Opt. Soc. Amer.*, **37**, 176.

Cartwright, C. H., and Turner, A. F., 1939 a, *Phys. Rev.*, **55**, 595; 1939 b, *Ibid.*, **55**, 1128.

Coleman, H. S., Clark, D. G., and Coleman, M. F., 1947, *J. Opt. Soc. Amer.*, **37**, 671.

Cotton, P., 1947, *Ann. Phys., Paris*, **2**, 209.

Craven, E. C., 1945, *Proc. Phys. Soc.*, **57**, 97.

Crawford, M. F., Gray, W. M., Schawlow, A. L., and Kelly, F. M., 1949, *J. Opt. Soc. Amer.*, **39**, 888.

* This full publication of the papers read at the Conference in Marseille appeared after the manuscript of this report had been completed. Individual references are therefore not quoted.

CROOK, A. W., 1948, *J. Opt. Soc. Amer.*, **38**, 954.
DUFFIEUX, M., 1932, *Rev. Opt. (théor. instrum.)*, **11**, 159.
DUFOUR, CH., 1945, *Rev. Opt. (théor. instrum.)*, **24**, 11; 1948, *Le Vide*, **3**, 480.
DUFOUR, CH., and PICCA, R., 1945, *Rev. Opt. (théor. instrum.)*, **24**, 19.
DYSON, J., 1949 a, *Proc. Phys. Soc.* B, **62**, 565; 1949 b, *Nature, Lond.*, **164**, 229.
EDWARDS, H. W., 1933, *Phys. Rev.*, **43**, 205.
FABRY, CH., and PEROT, A., 1902, *Astrophys. J.*, **15**, 73, 261; 1905, *C. R. Acad. Sci., Paris*, **140**, 848.
FAUST, R. C., and TOLANSKY, S., 1947, *Proc. Phys. Soc.*, **59**, 951.
FREDERIKSE, A. M., 1933, *Acta brev. Neerl. Physiol.*, **3**, 8.
GAYHART, E. L., and PRESCOTT, R., 1949, *J. Opt. Soc. Amer.*, **39**, 564.
GEE, A. E., and POLSTER, H. D., 1949, *J. Opt. Soc. Amer.*, **39**, 1044.
GEFFCKEN, W., 1942, German Patent No. 716153; 1948, *Angew. Chem.* A, **60**, 1; 1949, *Ceram. Abstr.*, **32**, 87.
GEHRCKE, E., and LAU, E., 1927, *Z. techn. Phys.*, **4**, 157; 1931, *Ann. Phys., Lpz.*, **10**, 71.
GOOS, F., 1936, *Z. Phys.*, **100**, 95.
GREENLAND, K. M., 1943, *Nature, Lond.*, **152**, 290; 1946, *J. Sci. Instrum.*, **23**, 48.
GREENLAND, K. M., and BILLINGTON, C., 1950, *Proc. Phys. Soc.* B, **63**, 359.
HADLEY, L. N., and DENNISON, D. M., 1946, *Phys. Rev.*, **69**, 258; 1947, *J. Opt. Soc. Amer.*, **37**, 451; 1948, *Ibid.*, **38**, 483.
HAMMER, K., and v. MILLER, O., 1949, *Optik*, **5**, 365.
HAMY, M., 1906, *J. Phys. Radium*, **5**, 789.
HIESINGER, L., 1948, *Optik*, **3**, 485.
HOLDEN, J., 1949, *Proc. Phys. Soc.* B, **62**, 405.
HOUSTON, W. V., 1927, *Phys. Rev.*, **29**, 478.
JACKSON, D. A., and KUHN, H., 1938, *Proc. Roy. Soc.* A, **167**, 205; 1939, *Ibid.*, **173**, 278.
JACQUINOT, P., 1942, *Rev. Opt. (théor. instrum.)*, **21**, 15.
JACQUINOT, P., and DUFOUR, CH., 1949, *J. Rech. Cent. Nat. Rech. Sci.* (No. 6), 91.
JOHNSON, E. R., and SCHOLES, J. F. M., 1948, *Aust. J. Sci. Res.*, A **1**, 464.
KINDER, W., 1946, *Optik*, **1**, 413.
KING, P., and LOCKHART, L. B., 1946, *J. Opt. Soc. Amer.*, **36**, 513.
KOCH, J., 1948, *Nature, Lond.*, **161**, 566.
KÖSTERS, W., 1926, *Z. Feinmech.*, **34**, 55.
KÖSTERS, W., and LAMPE, P., 1934, *Phys. Z.*, **35**, 223.
KRAUTKRÄMER, J., 1938, *Ann. Phys., Lpz.*, **32**, 537.
DE KRONIG, R., BLAISSE, B. S., and v. D. SANDE, J. J., 1938, *Appl. Sci. Res.* B, **1**, 63.
KUHN, H., and WHEATLEY, G. A., 1945, *J. Sci. Instrum.*, **22**, 110.
KUHN, H., and WILSON, B. A., 1950, *Proc. Phys. Soc.* B, **63**, 745.
LADENBURG, R., and BERSHADER, D., 1949, *Rev. Mod. Phys.*, **21**, 510.
LADENBURG, R., WINCKLER, J., and VAN VOORHIS, C. C., 1948, *Phys. Rev.*, **73**, 1359; 1949, *Ibid.*, **76**, 662.
LÆURGANS, P., and TURNER, A. F., 1947, *J. Opt. Soc. Amer.*, **37**, 983.
LITTLEFIELD, T. A., 1946, *Proc. Roy. Soc.* A, **187**, 220; 1950, *Nature, Lond.*, **165**, 187.
LOCKHART, L. B., and KING, P., 1947, *J. Opt. Soc. Amer.*, **37**, 689.
LOTMAR, W., 1949, *Helv. Chim. Acta*, **32**, 1847.
LYOT, B., 1933, *C. R. Acad. Sci., Paris*, **197**, 1593; 1944, *Ann. Astrophys.*, **7**, No. 1–2.
MACLAURIN, R. C., 1906, *Proc. Roy. Soc.* A, **78**, 296.
MCNALLY, J. R., GRIFFIN, P. M., and BURKHART, L. E., 1949, *J. Opt. Soc. Amer.*, **39**, 1036.
MALÉ, D., 1945, *J. Phys. Radium*, **12**, 40.
MARCELIN, A., 1931, *J. Chim. Phys.*, **28**, 605.
MEGGERS, W. F., 1948, *J. Opt. Soc. Amer.*, **38**, 7.
MEISSNER, K. W., 1942, *J. Opt. Soc. Amer.*, **32**, 185.
MERTON, T., 1947 a, *Proc. Roy. Soc.* A, **189**, 309; 1947 b, *Ibid.*, **191**, 1.
MESSNER, R., 1943, *Zeiss Nachr.*, **4**, 253; 1947, *Optik*, **2**, 228.
MOONEY, R. L., 1945, *J. Opt. Soc. Amer.*, **35**, 574; 1946, *Ibid.*, **36**, 254.
MUCHMORE, R. B., 1948, *J. Opt. Soc. Amer.*, **38**, 20.
MURMANN, H., 1933, *Z. Phys.*, **80**, 161.
PECK, E. R., 1948, *J. Opt. Soc. Amer.*, **38**, 66 and 1015.
PÉRARD, A., 1934, *Trav. Bur. Int. Poids Mes.*, **19**, 66.

Perrot, M., 1943, *Rev. Opt. (théor. instrum.)*, **22**, 25; 1949, *Ibid.*, **28**, 564.
Philpot, J. St. L., 1948, Brit. Pat. Appl. No. 15260/48.
Polster, H. D., 1949, *J. Opt. Soc. Amer.*, **39**, 1038.
Pringsheim, E., 1898, *Verh. dtsch. Phys. Ges.*, **17**, 152.
Ramachandran, G. N., 1942, *Proc. Indian Acad. Sci.* A, **16**, 336.
Räntsch, K., 1949, *Optik*, **4**, 283.
Rasmussen, E., 1945, *K. Danske Vidensk. Selsk. Mat.-fys. Medd.*, **23**, 1.
Rayleigh, Lord, 1888, *Phil. Mag.*, **26**, 256.
Rood, J. L. 1949, *J. Opt. Soc. Amer.*, **39**, 854.
Rouard, P., 1932, *C. R. Acad. Sci., Paris*, **195**, 869; 1949, *J. Phys. Radium* **10**, 120.
Sennett, R. S., and Scott, G. D., 1950, *J. Opt. Soc. Amer.*, **40**, 203.
Series, G. W., 1950, *Nature, Lond.*, **166**, 961.
Shane, C. D., 1932, *Lick Obs. Bull.*, **16**, 76; 1941, *Ibid.*, **19**, 119.
Sirks, J. A., 1893, *Handl. Ned. Nat. en Gen. Congr. Groningen.*
Sittner, W. R., and Peck, E. R., 1949, *J. Opt. Soc. Amer.*, **39**, 544.
Strong, J., 1936, *J. Opt. Soc. Amer.*, **26**, 73.
Strong, J., and Dibble, B., 1940, *J. Opt. Soc. Amer.*, **30**, 431.
Subrahmanian, R. V., 1941, *Proc. Indian Acad. Sci.* A, **13**, 467.
Timms, C., 1945, *J. Sci. Instrum.* **22**, 245.
Tolansky, S., 1943, *Nature, Lond.*, **152**, 722; 1944 a, *Ibid.*, **153**, 195; 1944 b, *Phil. Mag.*, **35**, 120, 175; 1945 a, *Proc. Roy. Soc.* A, **184**, 41, 51; 1945 b, *Phil. Mag.*, **36**, 225; 1945 c, *J. Sci. Instrum.*, **22**, 161; 1946 a, *Proc. Roy. Soc.* A, **186**, 261; 1946 b, *Proc. Phys. Soc.*, **58**, 654; 1946 c, *Phil. Mag.*, **37**, 390, 453.
Tolansky, S., and Bardsley, W., 1948, *Nature, Lond.*, **161**, 925.
Tolansky, S., and Khamsavi, A., 1946, *Nature, Lond.*, **157**, 661; 1946, *Ibid.*, **158**, 519.
Tolansky, S., and Ranade, J., 1949, *Mon. Not. R. Astr. Soc.*, **109**, 1.
Tolansky, S., and Wilcock, W. L., 1947, *Proc. Roy. Soc.* A, **191**, 182.
Treanor, P. J., 1949, *Mon. Not. R. Astr. Soc.*, **109**, 37.
Turner, A. F., 1946, *J. Opt. Soc. Amer.*, **36**, 711.
Twyman, F., and Green, A., 1918, *Phil. Mag.*, **35**, 49.
Vasicek, A., 1947, *J. Opt. Soc. Amer.*, **37**, 623; 1949, *Ibid.*, **39**, 409.
Weber, A. P., 1928, *Phys. Z.*, **29**, 233.
Weinstein, W., 1947, *J. Opt. Soc. Amer.*, **37**, 576.
Wiens, J. H., 1946, *Phys. Rev.*, **70**, 910.
Wiens, J. H., and Alvarez, L. W., 1940, *Phys. Rev.*, **58**, 1005.
Wilcock, W. L., 1949, *J. Opt. Soc. Amer.*, **39**, 546.
Winckler, J., 1948, *Rev. Sci. Instrum.*, **19**, 307.
Wolter, H., 1937, *Z. Phys.*, **105**, 269.

THE DIFFRACTION THEORY OF ABERRATIONS

By E. WOLF*
The Observatory, Cambridge University

CONTENTS

§ 1. Introduction 95
§ 2. Geometrical treatment of aberrations 96
§ 3. The diffraction theory of aberrations 97
3.1. A historical review 97
3.2. Advances since 1940 101
3.21. Images in the presence of small aberrations 101
3.22. Tolerance conditions 106
3.23. Asymptotic behaviour of the diffraction integral 108
3.24. Images formed by waves of non-uniform amplitude 109
3.25. Other researches 112
Acknowledgments 119
References 119

ABSTRACT. A historical and critical survey is given of investigations concerned with image formation in optical instruments in the presence of aberrations. The development of the subject is traced from the early researches of Airy on an aberration-free image. The advances made in recent years are discussed in greater detail; these mainly concern the effects of small aberrations, tolerance criteria, and the asymptotic treatment of diffraction problems. A section is also included on investigations into the effects of waves of non-uniform amplitude. The detailed light distribution in typical images is illustrated by isophote-diagrams and photographs.

§ 1. INTRODUCTION

In a perfect optical system the light waves which proceed from each element of the object emerge in the image space as convergent waves spherical in form. Such an instrument represents an idealization which cannot be realized in practice; in the image space of an actual instrument, the wave surfaces will, as a rule, be of a more complicated form. The deviations from the ideal spherical form of these surfaces may range in practice from a fraction of a wavelength in a well corrected telescope or microscope objective, to several dozen wavelengths in instruments required for less precise work. In many cases the resulting image bears little resemblance either to the Airy diffraction pattern or to the confusion figure predicted by geometrical optics.

The diffraction theory of aberrations is concerned with the study of images formed by actual optical instruments. Recent years have seen important advances in this branch of optics and many valuable results have been obtained. In particular, in the domain of very small aberrations (amounting only to a fraction of a wavelength) greatly simplified series expansions for the light distribution in diffraction patterns have been given, supplemented by a detailed study of a number of typical cases. The effects of aberrations of intermediate size (about one to ten wavelengths) have been studied with the help of a specially constructed mechanical integrator. Attempts have also been made to examine the effects of large aberrations. Further, a new and systematic investigation of the maximum amount of aberrations which may be tolerated in optical instruments has been carried out. The possibility has also been examined of improving the quality of

* Now at the Department of Mathematical Physics, University of Edinburgh.

the image by deliberately introducing non-uniformity of amplitude and sometimes, in addition, varying the phase of the disturbance over the exit pupil.

As no article on the diffraction theory of aberrations has been published in the preceding volumes of these Reports we have included in the present account a historical review of the subject. The period before 1940 is, however, treated only briefly; additional information can be found in the articles of von Laue (1928), König (1929), Martin (1946) and in a thesis by Nijboer (1942). A survey of experimental investigations on diffraction images was recently given in a thesis by Nienhuis (1948). We begin with a short preliminary section on the geometrical treatment of aberrations; this serves as an introduction to the classification of image errors used in the main discussion.

§2. GEOMETRICAL TREATMENT OF ABERRATIONS

The geometrical theory of aberrations is usually developed by methods due to Hamilton based on Fermat's principle of stationary path. This principle asserts that if light is propagated from a point P′ to a point P in a medium of which the refractive index at a typical point is denoted by μ, it will travel along a path for which the optical length

$$\int_{P'}^{P} \mu \, ds \qquad \ldots\ldots(2.1)$$

is stationary with respect to a small variation of that path. The integral (2.1) when taken along a natural ray may be regarded as a function of six variables, namely the rectangular Cartesian coordinates of P′ and P with respect to some fixed reference system. In consequence of Fermat's principle, this function, sometimes called the characteristic function of the medium and denoted by $V(x', y', z'; x, y, z)$ satisfies the relations

$$\left.\begin{aligned} &\frac{\partial V}{\partial x'} = -\mu' l', \quad \frac{\partial V}{\partial y'} = -\mu' m', \quad \frac{\partial V}{\partial z'} = -\mu' n', \\ &\frac{\partial V}{\partial x} = \mu l, \qquad \frac{\partial V}{\partial y} = \mu m, \qquad \frac{\partial V}{\partial z} = \mu n. \end{aligned}\right\} \qquad \ldots\ldots(2.2)$$

Here (l', m', n') and (l, m, n) denote the direction cosines of the rays at P′ and at P respectively, while μ' and μ are the refractive indices at these points.

Consider an optical instrument with a point source at P_0' (x_0', y_0', z_0') emitting monochromatic light. From (2.2) it follows that the equations

$$V(x_0', y_0', z_0'; x, y, z) = \text{constant} \qquad \ldots\ldots(2.3)$$

represent the orthogonal trajectories of the rays, i.e. the wave fronts of geometrical optics. Except in the ideal case when (2.3) represents concentric spherical surfaces in the image space of the system, the rays will intersect the image plane at different points. Geometrical optics identifies the light intensity at a typical element of the image plane with the density of these ray intersections. Apart from the details of structure, this gives an approximation to the actual light distribution when the deformations of the wave fronts are of a considerable amount.

To investigate the appearance of the image it is convenient to introduce an aberration function Φ which measures the deviation of the wave fronts from the spherical form. We consider a system with cylindrical symmetry and take a reference sphere centred on the Gaussian image P_0 of P_0', passing through the

centre of O of the exit pupil. Further we denote by A a typical point on this sphere and by B the intersection of P_0 A with the wave which passes through O (Figure 1).

The aberration function, defined by the equation

$$\Phi = AB, \qquad \ldots\ldots(2.4)$$

may be expressed as function of the three rotational invariants σ^2, r^2 and $\sigma r\cos(\chi-\theta)$, where (r,θ), and (σ,χ) denote polar coordinates in the exit pupil and of the image point P_0 respectively. It may as a rule be expanded in the form *

$$\Phi(\sigma,r,\phi) = \sum_{i,j,k} a_{ijk}\sigma^{2i+k}r^j\cos^k\phi, \qquad \ldots\ldots(2.5)$$

where $\phi=\chi-\theta$, i,j,k are non-negative integers, $j\geqslant k$, $j-k$ is even ($\neq 0$), $2i+j+k\geqslant 4$, and the a's are constants.

Let P be the actual point of intersection with the image plane of a typical ray proceeding from the exit pupil. From Φ it is possible to determine with the help of Fermat's principle the ray aberration displacements P_0P. Those terms in (2.5) for which $2i+j+k=N$ are called the wave aberrations of the Nth order

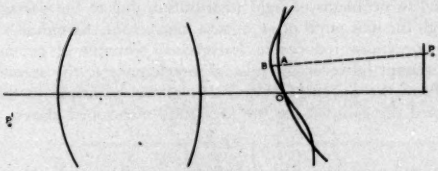


Figure 1.
The line BAP_0 does not necessarily lie in the plane $P_0'OP_0$. O is the centre of the exit pupil.

(N is always even) and give rise to ray aberrations of order † $N-1$. The five aberrations of the lowest order ($N=4$) are as a rule dominant in the paraxial region. The corresponding ray aberrations are known respectively as third order (sometimes primary or Seidel) spherical aberrations, coma, astigmatism, field curvature and distortion and have been discussed in the literature in much detail. Their physical meaning is too well known to need repeating here. For detailed discussions of geometrical aberrations we refer the reader to papers by Schwarzschild (1905), by Steward (1926) and by Nijboer (1943).

§ 3. THE DIFFRACTION THEORY OF ABERRATIONS

3.1. *A Historical Review*

We mentioned that the geometrical theory of aberrations identifies the intensity in the image plane with the ray density. This approximation, which in many cases gives an adequate picture of the light distribution in the image, gradually loses its validity as the aberrations become smaller. In the limiting case of perfectly spherical waves, for example, issuing from a circular opening,

* A different expansion recently proposed by Nijboer (1942, 1943) will be discussed later. The present notation is substantially due to him.

† A certain confusion exists in the literature about the terminology. Some earlier authors call first order ray aberrations those which we have denoted as ray aberrations of the third order. The classification used here has the advantage that it can be generalized to cover errors of focusing which are represented by first order terms. Such errors are often associated with the chromatic aberration.

geometrical optics predicts in the focal plane an infinite intensity at focus and zero intensity elsewhere; in reality the image consists of a bright central patch surrounded by rings—the familiar Airy pattern. It is therefore clear that in certain cases more refined investigations are needed. A direct application of electromagnetic theory presents considerable mathematical difficulties; instead one often uses approximations based on Kirchhoff's formula

$$U(P)=\frac{1}{4\pi}\iint_{D}\left\{U\frac{\partial}{\partial n}\left(\frac{e^{-ikr}}{r}\right)-\frac{e^{-ikr}}{r}\frac{\partial U}{\partial n}\right\}dS. \qquad \ldots\ldots(3.1)$$

This expresses the wave disturbance at any point P inside a closed region D as a surface integral involving the disturbance U and its gradient $\partial U/\partial n$ over the boundary. Here r represents the distance of a typical surface element from P, $\partial/\partial n$ denotes differentiation along the inward normal to D, $k=2\pi/\lambda$ and λ is the wavelength. In the investigations which we shall discuss, the scalar disturbance U is identified with a component of the electric or the magnetic vector or of a Hertz vector. From it the intensity I is derived by means of the approximate relation

$$I=\text{constant}\,|U|^2. \qquad \ldots\ldots(3.2)$$

When applied to problems of light distribution due to the passage of aberrant waves through the exit pupil of an optical instrument, Kirchhoff's formula with certain approximations reduces to Huyghens' principle of geometrical optics, taking into account however the mutual interference of the secondary wavelets proceeding from the opening. Using Huyghens' principle in this form, Airy (1835) obtained the solution for the ideal case mentioned above. It is usually written as

$$I(v)=\left[\frac{2J_1(v)}{v}\right]^2, \qquad \ldots\ldots(3.3)$$

where in suitable units $I(v)$ is the intensity in the pattern at distance v from its centre and J_1 is a Bessel function of the first kind.*

In two important papers Lommel (1885, 1886) extended Airy's classical result by deriving expressions for the intensity distribution due to spherical waves in planes other than the geometrical focal plane. For diffraction at a circular opening his solution takes the form†

$$I(u,v)=\left(\frac{2}{u}\right)^2[U_1^2(u,v)+U_2^2(u,v)], \qquad \ldots\ldots(3.4)$$

where the U's are two of the functions

$$U_n(u,v)=\sum_{s=0}^{\infty}(-1)^s\left(\frac{u}{v}\right)^{n+2s}J_{n+2s}(v), \qquad \ldots\ldots(3.5)$$

u is a parameter which specifies the position of the receiving plane and v has the same meaning as before. When $|u/v|>1$ the convergence of (3.5) is slow and for this case Lommel expressed the solution in terms of the functions

$$V_n(u,v)=\sum_{s=0}^{\infty}(-1)^s\left(\frac{v}{u}\right)^{n+2s}J_{n+2s}(v), \qquad \ldots\ldots(3.6)$$

* The derivation of (3.3) involved the assumption that the angular semi-aperture of the opening is small and that the disturbance over the converging wave is of uniform amplitude. In practice these assumptions may be far from valid but it has been shown by Hopkins (1943, 1944) that with a suitable interpretation of the argument v, Airy's formula is accurate to within a few per cent for systems with angular semi-aperture up to about 30°.

† The case of a circular opening was treated in an almost identical manner by Struve (1886). Formulae more convenient for calculating the intensity far from focus were given by Schwarzschild 1897).

which are related to the U_n functions by means of the relations

$$\left.\begin{aligned} U_{2n+1}(u,v)+V_{-2n+1}(u,v)&=(-1)^n\sin\frac{1}{2}\left(u+\frac{v^2}{u}\right)\\ -U_{2n+2}(u,v)+V_{-2n}(u,v)&=(-1)^n\cos\frac{1}{2}\left(u+\frac{v^2}{u}\right).\end{aligned}\right\}\quad\ldots\ldots(3.7)$$

The functions U_n and V_n which Lommel introduced and which now bear his name play an important part in a number of related problems. Lommel evaluated numerically the solution for a large number of particular cases. Full use of his results was however not made until forty years later when Berek (1926) obtained with their help a graphical representation of the three-dimensional light distribution near focus.* Recently, Zernike and Nijboer (1949) gave a more accurate and detailed diagram (see Figure 2) based on a different expansion of the diffraction integral.

The first investigations concerning diffraction images in the presence of monochromatic aberrations appear to be due to Rayleigh and Strehl. Rayleigh

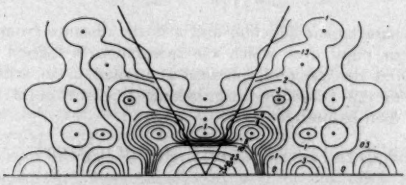


Figure 2. Isophotes (lines of equal intensity) near focus in a meridional plane in absence of aberrations. The straight lines indicate the boundary of the geometrical shadow. The numbers give intensity as percentage of intensity at focus. Axial minima and maxima are indicated by short strokes, others by small circles. *After* Zernike and Nijboer (1949).

(1879) studied images formed by cylindrical waves affected by a certain unsymmetrical aberration which are diffracted at a rectangular aperture, and also examined the effects of third order spherical aberration. In the latter case, however, he confined his investigations to determining the intensity at the centre of the pattern only. He formulated an important tolerance criterion, known in an extended form as Rayleigh's limit. This criterion asserts that the quality of an instrument is not sensibly affected by the presence of certain commonly occurring types of aberration if they are such that the waves in the image space do not deviate by more than a quarter of a wavelength from suitably chosen spherical surfaces. This criterion has since become widely used in formulating conditions concerning the maximum amount of aberrations which may be tolerated in the practical design of optical instruments.

Strehl, in his book *Théorie des Fernrohres* (1894) and in numerous papers mostly published between 1893 and 1930 in the *Zeitschrift für Instrumentkünde* and in *Zentralzeitung für Optik und Mechanik*, studied the effect of third order aberrations but he confined his researches mainly to investigating the variation of intensity along the principal ray. To him we owe the important

* Berek's figure is rather inaccurate. Later versions of it show the ouline of the geometrical shadow incorrectly as, for example, in the following books: *Handbuch der Physik*, 1929, 21, 885 and Picht, J., 1931, *Optische Abbildung* (Braunschweig: Vieweg), p. 71.

concept now known as Strehl definition (Definitionshelligkeit), viz. the ratio of the maximum intensity in a particular receiving plane of an instrument to the intensity at the centre of the Airy disc in a perfect system of the same aperture and focal length. This quantity supplies in many cases a good measure of the quality of the system.

Conrady (1919), Buxton (1921, 1923) and Martin (1922) investigated the effects of various aberrations with the help of numerical integrations. They computed a number of particular cases and obtained valuable information concerning Rayleigh's criterion. Some work of these authors was also concerned with balancing spherical aberrations of different order against each other so as to obtain high intensity at the centre of the pattern; this question was later studied in a more general manner by Richter (1925).

In important papers Steward (1925) and Picht (1925, 1926) derived series expansions for the intensity distribution in typical diffraction images. In Steward's treatment the diffraction integral, derived by an immediate application of Huyghens' principle, gives the intensity I at a point P of the pattern in the form

$$I = \left| \iint e^{ik(V+d)} dS \right|^2, \qquad \ldots\ldots (3.8)$$

where V is a characteristic function and d is the distance from P to a typical point of the exit pupil over which the integration is carried out. Steward considered at first the effect of spherical aberration in an arbitrary receiving plane parallel to the Gaussian image plane, the exit pupil being assumed circular. (3.8) then reduces to

$$I(u,v) = \left| \int_0^1 \exp(\tfrac{1}{2}iut + i \sum_{s\geq 2} A_s t^s) . J_0(v\sqrt{t})\, dt \right|^2. \qquad \ldots\ldots (3.9)$$

The problem treated earlier by Lommel corresponds to the case $A_s = 0$ for all s. By an argument similar to Lommel's, Steward showed that (3.9) can be expressed in terms of so-called generalized Lommel functions, these again being series involving Bessel functions.

When other aberrations are present (3.8) takes a more complicated form, since simplifying symmetry conditions no longer exist. Steward restricted the rest of his analysis to third order aberrations. His paper contains two diagrams showing isophotes (lines of equal intensity) in diffraction images, one affected by a certain amount of third order coma, the other by third order astigmatism. Unfortunately these diagrams are rather incomplete and contain some errors [see Nijboer (1947, p. 619) and Kingslake (1948, p. 152)]. Steward also studied the effects of different forms of aperture on the resolution. In particular he examined the influence of an annular aperture when third order spherical aberration is present under out-of-focus conditions.

Picht (1925, 1926) took as the starting point of his researches a well-known result due to Debye that the effect of spherical waves emerging from an optical instrument may be represented by a superposition of plane waves with different phases, amplitudes and directions of propagation. By generalizing Debye's formulae, Picht showed that the intensity distribution associated with the passage of aberrant waves through an exit pupil may be written in the form

$$I(x,y,z) = \left| \iint \psi(m,n) \exp\{-ik[(x-\xi)l + (y-\eta)m + (z-\zeta)n]\} \frac{dm\, dn}{l} \right|^2$$

where $\xi = \xi(m,n), \quad \eta = \eta(m,n), \quad \zeta = \zeta(m,n) \qquad \ldots\ldots (3.10)$

are parametric equations of a typical wave, (l, m, n) are direction cosines of its normals and ψ is an amplitude factor. In the presence of third order aberrations, Picht obtained the development of (3.10) into series of Bessel functions. From these series he computed, for a small amount of spherical aberration, the three-dimensional intensity distribution in the neighbourhood of the paraxial focus. He also studied a number of images affected by third order astigmatism in systems with a rectangular exit pupil.

Born (1932, 1938) studied the influence of very small aberrations of the third order. He derived an expression for the intensity, applicable whether one or more observations are present, in terms of the functions

$$K_0(v) = \left[\frac{J_1(v)}{v}\right]^2, \qquad K_1(v) = \frac{2J_1(v)J_2(v)}{v}, \qquad K_2(v) = \frac{2J_1(v)J_3(v)}{v^2}.$$

Born's formula applies however to distributions in the Gaussian image plane only.

No further substantial advances were made up to about 1940. The following years witnessed a renewed interest in this subject and led to a number of important developments which we shall now discuss.

3.2. *Advances since* 1940

3.21 *Images in the presence of small aberrations.*

Some of the researches so far described led to results which in many cases permitted the calculation of the light distribution in the diffraction images. These solutions, however, were not entirely satisfactory as they were either too restricted or involved very heavy computations.

In an important thesis * Nijboer (1942) took up this problem afresh and obtained a simpler and more satisfactory solution for cases where the wave deformation is small, only a fraction of a wavelength. Nijboer considered two series expansions of the aberration function Φ, viz.

$$\Phi(\sigma, r, \phi) = \sum_{l,m,n} b'_{lnm}\sigma^{2l+m} r^n \cos m\phi \qquad \ldots\ldots(3.11)$$

and

$$\Phi(\sigma, r, \phi) = \sum_{l,m,n} b_{lnm}\sigma^{2l+m} R_n^m(r) \cos m\phi, \qquad \ldots\ldots(3.12)$$

where l, m, n are non-negative integers, $n \geqslant m$, $n-m$ is even, $R_n^m(r)$ are certain polynomials which we discuss later, the other symbols having same meaning as in §2. A closer examination shows that the first expansion is particularly suitable for a geometrical treatment of aberrations, the second for a diffraction treatment. We recall that the traditional expansion with the same variables would contain terms $\cos^m\phi$ in place of $\cos m\phi$. The occurrence of the Fourier terms is essential in Nijboer's theory.

Some advantages of the new expansions can be illustrated by considering first the geometrical aberrations. It can be shown that in the presence of a single aberration

$$b'_{lnm}\sigma^{2l+m} r^n \cos m\phi^m \qquad \ldots\ldots(3.13)$$

the rays from the point (r, ϕ) of the exit pupil intersect the plane through the Gaussian image point perpendicular to the principal ray in a point (x, y) given by

$$x + iy = \tfrac{1}{2} b'_{lnm}\sigma^{2l+m} r^{n-1} \frac{R}{a}\{(n+m)e^{-i(m-1)\varphi} + (n-m)e^{i(m+1)\varphi}\}, \qquad \ldots\ldots(3.14)$$

* An account of this work will also be found in Nijboer (1943, 1947). As mentioned by Nijboer the method was to some extent suggested by earlier unpublished researches of Zernike.

where a and R denote the radius of the exit pupil and of the reference sphere respectively. From (3.14) it follows that the main features of the aberration figure (given by $r=$constant), such as symmetry, depend on m while details depend both on m and n. It will be shown further that m and n play a similar role in the appearance of the diffraction pattern for an aberration specified by the term

$$b_{lnm}\sigma^{2l+m}R_n^m(r)\cos m\phi. \qquad \ldots\ldots(3.15)$$

This suggests a new classification of image errors in which the value of m determines the general type. By analogy with Seidel aberrations, the terms with $m=0$, 1 and 2 may be called spherical aberration, coma and astigmatism respectively. Curvature and distortion now appear as degenerate cases of spherical aberration and of coma. The terms with $m \geqslant 3$ do not occur in Seidel theory. This classification considerably simplies the discussion of the higher order effects.

In Nijboer's diffraction treatment the expansion of the aberration function is of the form (3.12) where

$$R_n^m(r)\cos m\phi \qquad \ldots\ldots(3.16)$$

are so-called circle polynomials introduced by Zernike (1934) in his diffraction theory of phase contrast. These polynomials are orthogonal within the unit circle and the $R_n^m(r)$ are given by

$$R_n^m(r)=(-1)^{(n-m)/2}\binom{(n+m)/2}{m}r^m F[\tfrac{1}{2}(n+m+2),-\tfrac{1}{2}(n+m),-\tfrac{1}{2}(n-m),m+1,r^2]$$

$$=\sum_{s=0}^{(n-m)/2}(-1)^s\frac{(n-s)!}{s!\,\{\frac{1}{2}(n+m)-s\}!\,\{\frac{1}{2}(n-m)-s\}!}r^{n-2s}, \qquad \ldots\ldots(3.17)$$

where F is a hypergeometric function. The expansion in terms of circle polynomials is particularly advantageous for the balancing of very small aberrations against each other so as to obtain maximum Strehl definition. We observe that in (3.15) a number of terms of the customary form $\sigma^{2l+m}r^j\cos^j\phi$ are combined; with the help of the orthogonality relations for the circle polynomials it may be shown that they have been combined in such a way as to give maximum intensity in the centre of coordinates for a given value of the coefficient of

$$\sigma^{2l+m}r^n\cos^m\phi.$$

This may be illustrated by recalling that Richter (1925) showed (in a different notation) that in the presence of a small amount of spherical aberration of the form

$$\Phi=A_6r^6+A_4r^4+A_2r^2+A_0, \qquad \ldots\ldots(3.18)$$

the Strehl definition is a maximum for a given value of A_6 if

$$A_4/A_6=-3/2, \quad A_2/A_6=3/5. \qquad \ldots\ldots(3.19)$$

In Nijboer's diffraction treatment the fifth order spherical aberration is given by the term

$$b_{060}R_6^0(r)=b_{060}(20r^6-30r^4+12r^2-1), \qquad \ldots\ldots(3.20)$$

and Richter's condition is automatically satisfied. However, we must insist that this balancing of aberrations as achieved in Nijboer's theory holds for very small aberrations only; in some cases, as Figure 8 indicates, it no longer holds when the maximum wave deformation is only about $0{\cdot}6\lambda$.

With the usual approximations the Huyghens–Kirchhoff integral for the disturbance at a point (q, ψ) in a receiving plane perpendicular to the principal ray and specified by a parameter p, can be written in the form

$$U(p,q,\psi)=\frac{1}{\pi}\int_0^1\int_0^{2\pi}\exp\{ipr^2+iqr\cos(\phi-\psi)-ik\Phi(\sigma,r,\phi)\}r\,dr\,d\phi. \quad \ldots\ldots(3.21)$$

In the presence of a single aberration Nijboer writes

$$\Phi(\sigma,r,\phi)=\frac{1}{k}\beta_{lnm}R_n^m(r)\cos m\phi,$$

where

$$\beta_{lnm}=k\sigma^{2l+m}b_{lnm} \quad \ldots\ldots(3.22)$$

and expands the term $\exp ik\Phi$ in powers of β. Neglecting fifth and higher powers of β in (3.21) and integrating with respect to ϕ Nijboer obtains

$$\begin{aligned}U(p,q,\psi)=&\,2\int_0^1 e^{ipr^2}J_0(qr)r\,dr-2i\beta_{lnm}i^m\cos m\psi\int_0^1 e^{ipr^2}R_n^m(r)J_m(qr)r\,dr\\
&+\frac{(i\beta_{lnm})^2}{2!}\left\{\int_0^1 e^{ipr^2}\{R_n^m(r)\}^2J_0(qr)r\,dr+i^{2m}\cos 2m\psi\right.\\
&\left.\times\int_0^1 e^{ipr^2}\{R_n^m(r)\}^2J_{2m}(qr)r\,dr\right\}-\frac{(i\beta_{lnm})^3}{2.3!}\left\{3i^m\cos m\psi\right.\\
&\times\int_0^1 e^{ipr^2}\{R_n^m(r)\}^3J_m(qr)r\,dr+i^{3m}\cos 3m\psi\int_0^1 e^{ipr^2}\{R_n^m(r)\}^3J_{3m}(qr)r\,dr\Big\}\\
&+\frac{(i\beta_{lnm})^4}{2^2.4!}\left\{3\int_0^1 e^{ipr^2}\{R_n^m(r)\}^4J_0(qr)r\,dr+4i^{2m}\cos 2m\psi\right.\\
&\times\int_0^1 e^{ipr^2}\{R_n^m(r)\}^4J_{2m}(qr)r\,dr+i^{4m}\cos 4m\psi\\
&\left.\times\int_0^1 e^{ipr^2}\{R_n^m(r)\}^4J_{4m}(qr)r\,dr\right\}. \quad \ldots\ldots(3.23)\end{aligned}$$

From this expression a number of symmetry properties of the diffraction pattern may immediately be deduced. It is seen for example that the p-axis is an m-fold axis of symmetry. Moreover if m is odd (e.g. as in the case of coma) the intensity distribution is symmetrical about the plane $p=0$. If m is even (but $m\neq 0$) the intensity at any point is equal to that in the point resulting from a reflection in the plane $p=0$ and the additional rotation about the p-axis through an angle π/m. For some special cases this rule has previously been stated by Strehl, by Picht and by Steward.

Nijboer outlined a method based on various properties of the circle polynomials which makes possible the development of the integrals in (3.23) into series of Bessel functions. The series terminate in the important case $p=0$. From these expansions the isophotes near focus may be determined for a small arbitrary single aberration. Some of the diagrams obtained in this way by Nijboer (1942, 1947) and Zernike and Nijboer (1949) are reproduced in Figures *3, 4, 5, 7. Additional cases of slightly larger aberrations were studied by Nienhuis and Nijboer (1949). This later investigation was confined to third order effects and contains new formulae and diagrams (Figures 6 and 8) for the

* Unless otherwise stated, all figures refer to systems with a circular exit pupil and $\Phi_{\max}$ denotes the maximum deviation of the wave from the reference sphere, shown in Figure 1. The Figures are not all reproduced on the same scale.

Figure 3. Isophotes as in Figure 2, in presence of third order spherical aberration of amount $\beta_{040}=\frac{1}{2}$ ($\Phi_{max}=0{\cdot}48\lambda$). The thick line indicates the geometrical caustic. Strehl definition 0·95. *After* Zernike and Nijboer (1949).

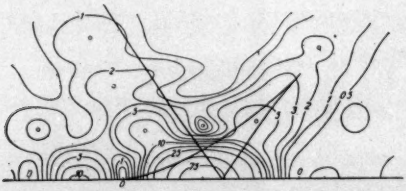


Figure 4. Isophotes as in Figure 2, in presence of spherical aberration of amount $\beta_{060}=\frac{1}{2}$ (1·6λ of fifth order spherical aberration balanced against 2·4λ of third order spherical aberration). The thick line indicates the geometrical caustic. Strehl definition 0·965. *After* Zernike and Nijboer (1949).

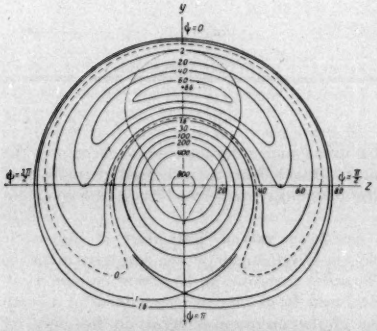


Figure 5. Isophotes in the Gaussian image plane in presence of third order coma of amount $\beta_{131}=1$ ($\Phi_{max}=0{\cdot}48\lambda$). Strehl definition 0·879. The dotted curve represents the line of zero intensity. The boundary of the geometrical confusion figure is also indicated. The numbers represent intensities corresponding to a value of 1,000 in the centre of the ideal Airy pattern. *After* Nijboer (1947).

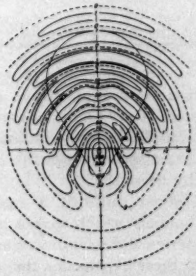


Figure 6. Isophotes in the Gaussian image plane in presence of third order coma of amount $\beta_{131}=3$ ($\Phi_{max}=1{\cdot}4\lambda$). Strehl definition 0·306. The dotted curves represent the line of zero intensity. The boundary of the geometrical confusion figure is also indicated. *After* Nienhuis and Nijboer (1949).

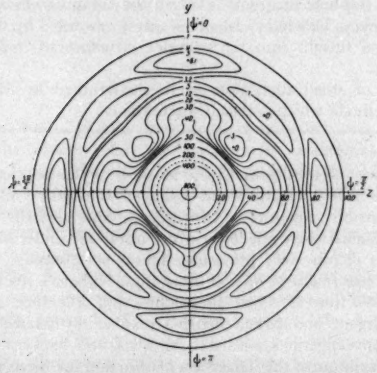


Figure 7. Isophotes in the central plane in presence of third order astigmatism of amount $\beta_{131}=1(\Phi_{max}=0{\cdot}16\lambda)$. Strehl definition 0·840. The dotted circle indicates the boundary of the geometrical confusion figure. *After* Nijboer (1947).

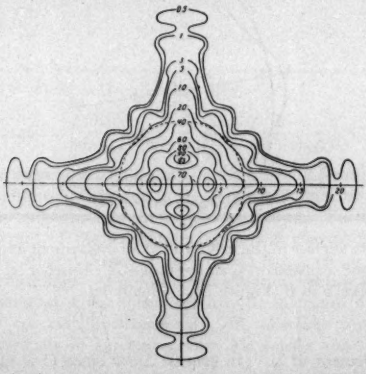


Figure 8. Isophotes in the central plane in presence of third order astigmatism of amount $\beta_{222}=4(\Phi_{max}=0{\cdot}64\lambda)$. Strehl definition 0·066. The dotted circle indicates the boundary of the geometrical confusion figure. *After* Nienhuis and Nijboer (1949).

light distribution in certain receiving planes. It was supplemented by photographs (see Plates *) of actual images showing a good agreement with the theoretical predictions. In this work the authors rediscover the formula

$$I=\left[\frac{2}{\beta}\sum_{s=0}^{\infty}J_{2s+1}(\beta)\right]^2,$$

originally given in a slightly different notation by Picht (1931, p. 202) for the Strehl definition in the central plane $p=0$ when third order astigmatism is present.

* For all Plates see end of volume.

Some of the isophote diagrams involved the computation of intensity at as many as 600 points. This heavy labour is amply rewarded by the results which give the first real insight into the complex structure of important types of diffraction image.

A few cases of small aberrations were also studied by Maréchal and by Kingslake by methods which we shall discuss later.

3.22. *Tolerance conditions.*

Wang Ta-Hang (1941) discussed the location of the best axial focus in the presence of spherical aberration and deduced tolerance conditions for the maximum permissible amounts of third, fifth and seventh order spherical aberration for systems where the lower order terms are under control. Most of Wang Ta-Hang's results were previously found by Conrady, Buxton, Martin and Richter and also follow as special cases from Nijboer's researches.

Optimum correction of small aberrations and tolerance conditions were studied in an elegant and comprehensive manner by Maréchal (1944 a, b, c, 1947 a). These investigations are based on a relation between the intensity at points near the maximum of the diffraction pattern and the mean square deviation of the wave front from a certain reference sphere.

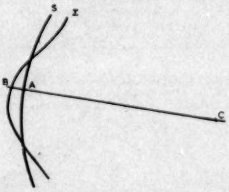


Figure 18.

Let Σ be a wave surface in the exit pupil of an instrument and let S be a sphere in its neighbourhood, centred on C and of radius R. Further let A be a point on S and let AC intersect Σ in B (Figure 18). Finally let

$$\mathrm{AB}=\Delta, \quad \mathrm{BC}=l, \quad d\omega = d\sigma \Big/ \iint_\Sigma d\sigma, \qquad \ldots\ldots(3.24)$$

where $d\sigma$ is an element of Σ. (In general $\Delta \neq \Phi$ since C is not necessarily the Gauss image point and S does not necessarily pass through the centre of the exit pupil.)

For a given wave front and for fixed C, the mean square deviation

$$E = \iint_\Sigma \Delta^2 d\omega, \qquad \ldots\ldots(3.25)$$

of the wave front from the sphere is a function of the radius R. This function has a minimum

when

$$\left.\begin{aligned} E = E_0 &= \iint_\Sigma l^2 d\omega - \left(\iint_\Sigma l d\omega\right)^2, \\ R = R_0 &= \iint_\Sigma l d\omega. \end{aligned}\right\} \qquad \ldots\ldots(3.26)$$

Maréchal showed, that when $|\Delta|<\lambda/4$ (a condition twice less severe than that imposed by Rayleigh's criterion), the intensity I (taken as unity at the centre of the Airy pattern) at C satisfies the relation *

$$I\sim\left[1-\frac{2\pi^2}{\lambda^2}E_0\right]^2. \qquad \ldots\ldots(3.27)$$

Earlier authors found that in presence of certain aberrations of amount $\frac{1}{4}\lambda$, the central intensity takes on values between about 0·73 and 0·87, the exact amount depending on the type of aberration. With the help of (3.27) Maréchal was able to formulate a new tolerance criterion which unlike that of Rayleigh corresponds to a fixed lower limit for the central intensity.

A system is generally regarded as well corrected if the loss in intensity at the centre of the pattern does not exceed 20%, i.e. when $I\geqslant 0{\cdot}8$. For such a case (3.27) implies that $E_0<\lambda^2/180$. Maréchal therefore concluded that for a well corrected system the root-mean-square deviation of the wave from the 'mean focal square' (given by $R=R_0$) is less than $\lambda/\sqrt{180}$. In accordance with this criterion, Maréchal then deduced the following tolerance conditions for defocusing and for aberrations of the third order:

Defocusing: $\lambda/2\alpha^2$.

Spherical aberration: $4\lambda/\alpha^2$.

Coma: $\left|\frac{\delta g}{g}\right|\leqslant 0{\cdot}63\,\frac{\lambda}{\sigma\alpha}$.

Astigmatism: $|t-s|\leqslant\lambda/\alpha^2\sqrt{2}$.

Astigmatism and curvature: $(t+s)^2+2(t-s)^2\leqslant\lambda^2/\alpha^4$.

Combined aberrations: $\left(\frac{L}{L_{\max}}\right)^2+\left(\frac{\delta g}{\delta g_{\max}}\right)^2+\left(\frac{t-s}{(t-s)_{\max}}\right)^2+\left(\frac{t+s}{(t+s)_{\max}}\right)^2\leqslant 1$.

Here α denotes the maximum angular semi-aperture in the image space, L denotes the longitudinal spherical aberration, g the sine condition ratio, and t and s are the abscissae of the tangential and the sagittal focal lines. Further, $L_{\max}$ denotes the maximum amount of the longitudinal spherical aberration which may be tolerated when no other aberration is present, $\delta g_{\max}$, $(t-s)_{\max}$ and $(t+s)_{\max}$ having analogous meanings. For fifth order spherical aberration and fifth order coma Maréchal found that the tolerances may be increased by 50% provided these aberrations are suitably 'balanced'. A tolerance condition for third order coma substantially in agreement with Maréchal's was recently deduced by Kingslake (1948).

Toraldo di Francia (1946) derived an expression for the lower limit of intensity at the centre of the reference sphere in terms of the maximum wave aberration. In particular, this expression shows that when the aberration is less than $\lambda/4$ the intensity at the centre does not fall below 0·5.

Françon (1944, 1947 a, b, 1948 a) studied both theoretically and experimentally the efficiency of visual optical instruments suffering from spherical aberration. (Aberrations may be tolerated when the efficiency, defined as the ratio of the limiting resolvable separation of the unaided eye to that of the instrument–eye combination, is nearly 1.) He found that for a small pupil (0·8–0·9 mm. in diameter), the eye may be regarded as a perfect instrument and in this case a

* Väisälä (1922) derived a weaker condition $I\sim 1-4\pi^2E_0/\lambda^2$.

high degree of correction is called for. For pupils of larger size poorer correction may be tolerated. Françon showed that for small pupils and objects of very low contrast (~0·3), aberrations of the order of $\lambda/16$ may be significant. Rayleigh's and Maréchal's criteria apply to a perfect eye, and contrast 1. Françon's investigations give, for the case of spherical aberration, an extension of Rayleigh's criterion for pupils up to 4 mm. and for a range of contrasts.

3.23. *Asymptotic behaviour of the diffraction integral.*

The effects of aberrations which are large compared with the wavelength may be investigated by examining the asymptotic behaviour of the diffraction integral for large values of $k=2\pi/\lambda$. Van Kampen (1949) studied this problem by using a two-dimensional analogue of the principle of stationary phase (cf. van der Corput (1948), p. 206).

The Huyghens–Kirchhoff diffraction integral which describes the complex displacement at a point P in the image space of an instrument may be written in the form

$$U(\mathrm{P})=\iint_{\mathrm{D}} kg(x,y)e^{ikf(x,y)}\,dx\,dy, \qquad \ldots\ldots(3.28)$$

where D is a domain bounded by a finite number of analytic curves C. Van Kampen found that, on account of the rapidity of fluctuation of the exponential term in (3.28), the value of $U(\mathrm{P})$ depends substantially on the behaviour of the integrand near a limited number of 'critical points' which may be of three kinds: (1) internal points of D at which f is stationary, i.e. where $\partial f/\partial x=\partial f/\partial y=0$; (2) boundary points at which f is stationary, i.e. where $\partial f/\partial s=0$, ds being element of C; (3) corner points or boundary points where two analytic curves join.

In the neighbourhood of a critical point of the first kind, the integrand of (3.28) can be expanded in the form

$$kge^{ikf}=k\exp ika_{00}\exp ik(a_{20}x^2+a_{11}xy+a_{02}y^2+\ldots).(b_{00}+b_{10}x+\ldots). \quad ..(3.29)$$

To find the corresponding contribution to the integral van Kampen neglects in the exponent third and higher powers of x and y and integrates term by term over the ranges $-\infty<x<\infty$, $-\infty<y<\infty$. The resulting expansion is a power series in k^{-1} beginning with a constant term. For the contribution of a critical point of the second kind, van Kampen obtains in an analogous manner a series with $k^{-1/2}$, $k^{-3/2}$, $k^{-5/2}$ and for the contribution of corner points, a power series in k^{-1}, in both these cases the constant term being absent. The total effect is then obtained by adding the contribution from each of the critical points.

The principal term of the final expansion is independent of the wavelength and comprises the constant terms of the expansions at critical points of the first kind. This part of the solution can be interpreted as a contribution of 'geometrical optics' taking, however, into account the mutual interference of the wave patches arriving at P from the immediate neighbourhood of critical points of the first kind.

The rest of van Kampen's paper deals with applications to third order aberrations in systems with a circular aperture. He then found for example that in the case of astigmatism points of the second kind give rise to an asteroid pattern; this pattern was also discussed by Nienhuis (1948).

Although van Kampen's investigations give some valuable results, his analysis cannot be regarded as satisfactory. For example, van Kampen integrates the expansions over infinite ranges, a procedure which introduces non-negligible

errors and in some cases even leads to divergent integrals. To overcome the divergence difficulty he uses the well known but mathematically unsatisfactory artifice of multiplying the integrand by a term $e^{-\sigma^2}$ and after integration sets $\sigma=0$. It appears that only the leading terms of van Kampen's expansion have any real claim to validity, and these themselves cannot be regarded as having been rigorously derived.

3.24. *Images formed by waves on non-uniform amplitude.**

When studying images which are formed by an instrument of the usual design it is as a rule permissible to assume that the waves are of uniform amplitude. Several authors have in recent years considered the possibility of improving the quality of the image by a suitable variation of the amplitude and sometimes also of the phase of the disturbance over the exit pupil. This may be realized in practice, for example by evaporating thin films of metallic or dielectric substances on to the lenses or by means of specially constructed filters.

With the usual approximations, it follows from Huyghens' principle that the complex displacement $U(x, y)$ in the receiving plane of a system imaging a point source may be written in the form of a double Fourier integral

$$U(x,y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}P(p,q)e^{2\pi i(px+qy)}\,dp\,dq, \qquad \ldots\ldots(3.30)$$

where $P(p, q)$ is the so-called pupil function (sometimes called transmission function), which describes, in terms of the optical direction cosines p and q in the image space, the complex displacement over a spherical reference surface centred at $x=y=0$. The integration is only formally carried out over an infinite domain since P is taken as zero when $p^2+q^2>\rho_m^2$, ρ_m being the maximum numerical aperture in the image space. In the case of rotational symmetry (3.30) may be written as

$$U(r)=2\pi\int_0^{\infty}P(\rho)J_0(2\pi r\rho)\rho\,d\rho, \qquad \ldots\ldots(3.31)$$

where ρ is the zonal numerical aperture in the image space and r is the distance from the optical axis in the receiving plane, measured in wavelengths. (The symbols r and P now denote different quantities from those in previous sections.)

Luneberg (1944) considered the case of rotational symmetry and studied the effects of variation of amplitude alone. He found that amongst all diffraction patterns of equal energy the highest central maximum is given by the normal pattern $P(\rho)=$constant. In particular, any pattern which gives improved resolution must therefore have a lower central maximum. By a variational argument Luneberg deduced an expression for the pupil function in terms of the radius r_0 of the first dark ring. It appears that for $r_0<0{\cdot}31(2\pi/\rho_0)$ (i.e. about a half of the radius r_A of the first dark ring in the Airy pattern) the resulting pattern has a rather low central maximum.

Luneberg also investigated how to chose $P(\rho)$ so that the fraction of the total illumination

$$L(r_0)=\frac{\int_0^{r_0}|U(r)|^2 r\,dr}{\int_0^{\infty}|U(r)|^2 r\,dr} \qquad \ldots\ldots(3.32)$$

* The investigations discussed in this section do not fall under the heading of diffraction theory of aberrations in the customary sense but are included since they are closely related to the problem of optimizing resolution by modifying the waves.

which reaches the circle of radius $r_0 < r_A$ is as large as possible. He showed that the corresponding pupil function is given as a solution of a certain integral equation. He also discussed the resolution of objects of periodic structure and found that in this case it is impossible to increase the resolving power by varying the amplitude only, but that the contrast of the image may be improved considerably by this means.

Couder (1944) discussed amplitude filters which consist of a plano-concave lens filled with an absorbing liquid. The corresponding pupil function is of the form $P(\rho) = 10^{-k\rho^2}$. It is perhaps worth noting that Straubel (1902, 1931) who appears to have been the first to study effects of non-uniformity of amplitude on resolution also considered pupil functions of this form. Couder found that with such a filter it is possible to redistribute the illumination in the diffraction pattern so as to facilitate the observation of double stars whose components differ widely in brightness. Amplitude filters were also discussed by Lansraux (1946, 1949).

Boughon, Dossier and Jacquinot (Boughon *et al.* 1946, Jacquinot *et al.* 1949) considered applications of amplitude variation to spectroscopy. For diffraction at a slit (3.31) and (3.32) are replaced by the equations

$$U(x) = \int_{-\infty}^{+\infty} P(x') e^{2\pi i x x'} \, dx', \qquad \ldots\ldots(3.33)$$

and

$$L(a) = \frac{\int_{-a}^{+a} [U(x)]^2 \, dx}{\int_{-\infty}^{+\infty} [U(x)]^2 \, dx}. \qquad \ldots\ldots(3.34)$$

The above authors used the analysis of trigonometrical polynomials to determine pupil functions which for a given value of the parameter a make the fraction of the total illumination $L(a)$ as large as possible. They also considered distributions for which $U(x)$ behaves asymptotically like x^{-m}, m being a positive integer. They showed that by the choice of a suitable trigonometric polynomial as pupil function the effect of diffraction fringes can be reduced, facilitating the detection of satellites to bright spectral lines.

Expansion of pupil functions into series of Hermitian polynomials were considered by Duffieux (1946 b). Expansions into series of Legendre polynomials were suggested by Slevogt (1949).

Lansraux (1947 a, b) discussed the evaluation of the diffraction integral (3.31) for cases where $P(\rho)$ can be expanded in a series of lambda functions

$$\Lambda_s(\rho) = [1 - (\rho/\rho_m)^2]^{s-1}. \qquad \ldots\ldots(3.35)$$

He showed that when

$$P(\rho) = \sum_{s=1}^{\infty} a_s \Lambda_s(\rho), \qquad \ldots\ldots(3.36)$$

where the a's are constants

$$U(\rho) = \pi \rho_m^2 \sum_{s=1}^{\infty} a_s 2^s (s-1)! \frac{J_s(2\pi\rho_m r)}{(2\pi\rho_m r)^s}. \qquad \ldots\ldots(3.37)$$

Lansraux asserts that for a small aberration (3.37) is rapidly convergent. That this is not necessarily so can be illustrated by considering the simplest aberration, namely defocusing, when (3.37) reduces to that solution of Lommel which involves the U_n series. These series converge rapidly only at points in the geometrical shadow (case $|u/v| < 1$ in §3.1 above). In the direct beam of light even when very near focus the series (3.37) is therefore not suitable for computations.

Hopkins (1949) extended Lommel's analysis to waves of non-uniform amplitude, the non-uniformity considered being of the form $\alpha+\beta(\rho/\rho_m)^2$ where α and β are constants. He obtained a solution in terms of Lommel's original U_n and V_n functions and certain closely related functions and gave a number of intensity distribution curves for out-of-focus patterns. He showed that for most ordinary lens systems it is quite justifiable in evaluating the diffraction integral to assume the disturbance over the converging waves to be of uniform amplitude.

Osterberg and Wilkins (1949) studied theoretically the problem of coating the exit pupil with light absorbing and refracting materials so as to reduce the diameter of the central bright disc in the diffraction pattern. They considered pupil functions of the form (3.36) and derived Lansraux' expression very shortly by the application of Sonine's integral formula. In some cases choice can be

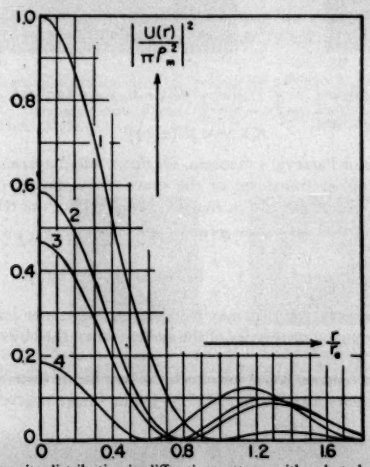


Figure 19. The intensity distribution in diffraction patterns with selected radii r of the first dark ring, the central maximum being the highest possible. Curves 2, 3, 4 correspond to $r=0{\cdot}8r_0$, $0{\cdot}75r_0$ and $0{\cdot}6r_0$ respectively, where r_0 denotes the radius of the first dark ring in the Airy pattern (curve 1). *After* Wilkins (1950).

made of the constants a_s to give pupil functions corresponding to diffraction patterns with prescribed properties. As Osterberg and Wilkins point out, however, it is not possible to go far in this direction because an arbitrary function $U(\rho)$ cannot, in general, be expressed in series of the type (3.37). The resolution of two particles in a bright field by Sonine type microscope objectives (microscope objectives for which $P(\rho)$ is given by (3.36)) was discussed in a paper by Osterberg and Wissler (1949).

Wilkins (1950) showed how to coat the exit pupil so as to obtain for any prescribed diameter of the central disc the highest possible central intensity. In Figure 19 we reproduce the intensity curves which he obtained for a few typical cases.

3.25. *Other researches.*

In his book *L'Intégrale de Fourier et ses Application a l'Optiques* and in a series of papers, Duffieux (1945 a, b, 1946 b, 1947, 1948) has, partly in collaboration with Lansraux (1945), attempted to apply the methods of Fourier transforms to the study of Fraunhofer diffraction.*

We recall that with a suitable interpretation of the variables p, q, x and y the complex disturbances $P(p,q)$ in the exit pupil and $U(x,y)$ in the 'pattern at infinity' due to a point source are related by the pair of double Fourier integrals

$$\left.\begin{aligned} U(x,y) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} P(p,q)e^{-2\pi i(px+qy)}\,dp\,dq, \\ P(p,q) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} V(x,y)e^{2\pi i(px+qy)}\,dx\,dy. \end{aligned}\right\} \quad \ldots\ldots(3.38)$$

If the system images an *extended incoherent object* in which the intensity distribution is given by $I(x,y)$, then the corresponding intensity distribution in the image is given by

$$\Pi(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} I(x',y')E(x-x',y-y')\,dx'\,dy', \quad \ldots\ldots(3.39)$$

where

$$E(x,y) = |U(x,y)|^2. \quad \ldots\ldots(3.40)$$

By the application of Parseval's theorem, Duffieux and Lansraux (1945) showed that the double Fourier transform or the space frequency spectrum $T[\Pi(x,y)]$ is connected with that of the object, namely $T[I(x,y)]$, by the relations

$$T[\Pi(x,y)] = T[I(x,y)]\,T[E(x,y)] = D(p,q)\,T[I(x,y)] \quad \ldots\ldots(3.41)$$

where

$$D(p,q) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} E(x',y')e^{2\pi i(px+qy)}\,dx\,dy. \quad \ldots\ldots(3.42)$$

From (3.41) it follows that $D(p,q)$ may be regarded as a transmission factor which characterizes the imaging properties of the system when the object is incoherently illuminated.

For an *extended coherent object* for which the amplitude distribution is denoted by $I'(x,y)$ the intensity distribution $\Pi'(x,y)$ in the image is given by

$$\Pi'(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} I'(x,y)U(x-x',y-y')\,dx'\,dy' \quad \ldots\ldots(3.43)$$

leading to a relation

$$T[\Pi'(x,y)] = T[I'(x,y)]\,T[U(x,y)] = P(p,q)\,T[I'(x,y)]. \quad \ldots\ldots(3.44)$$

(3.44) shows that $P(p,q)$ may be regarded as a transmission factor when the object is coherently illuminated.

From D and P Duffieux and Lansraux derived other transmission factors which have a bearing on the theory of test objects for optical instruments. To one such factor $d(p)$, defined by

$$d(p) = \int_{-\infty}^{+\infty} D(p,q)\,dq, \quad \ldots\ldots(3.45)$$

is related to resolving power in the image of non-coherent objects. By non-rigorous mathematical arguments Duffieux and Lansraux derived a number

* Additional references to papers dealing mainly with cases where the distribution in the image is a function of one variable only will be found in Duffieux (1946 a).

The existence of a Fourier integral relation between the disturbance over a spherical surface filling the exit pupil and the disturbance over its 'focal sphere' was first pointed out by Michelson (1905).

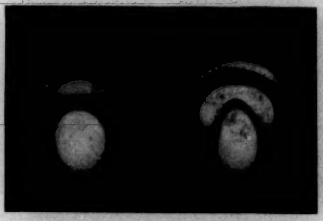


Figure 9. Images in Gaussian plane in presence of third order coma of amount $\Phi_{max}=0{\cdot}48\lambda$ and $\Phi_{max}=1{\cdot}4\lambda$. Compare with Figures 5 and 6. *After* Nienhuis (1948).

Figure 10. Image in Gaussian plane in presence of third order coma of amount $\Phi_{max}=7{\cdot}2\lambda$. Semicircular aperture. *After* Nienhuis (1948).

Figure 11. Images in central plane in presence of third order astigmatism with $\Phi_{max}=0{\cdot}16\lambda$ and with $\Phi_{max}=0{\cdot}64\lambda$. Compare with Figures 7 and 8. *After* Nienhuis (1948).

Figure 12. Images in presence of third order astigmatism of amount $\Phi_{max}=2{\cdot}7\lambda$ in central plane and in plane containing a focal line. *After* Nienhuis, *reprinted from* Zernike 1948.

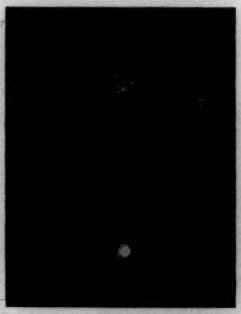


Figure 13. Images in presence of third order spherical aberration of amount $\Phi_{max}=16\lambda$ at marginal focus and at circle of least confusion. *After* Nienhuis (1948).

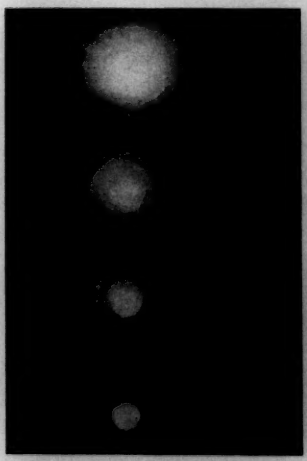


Figure 14. Images in plane of paraxial focus, in presence of third order spherical aberration of amounts $\Phi_{max}=17{\cdot}5\lambda$, $8{\cdot}4\lambda$, $3{\cdot}7\lambda$, $1{\cdot}4\lambda$. *After* Nienhuis (1948).

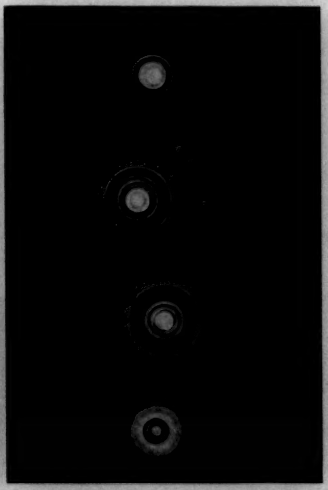


Figure 15. Images in plane of circle of least confusion in presence of third order spherical aberration of amounts $\Phi_{max}=17{\cdot}5\lambda$, $8{\cdot}4\lambda$, $3{\cdot}7\lambda$, $1{\cdot}4\lambda$. Scale $3\times$ as in Figure 14. *After* Nienhuis (1948).

Figure 16. Images in Gaussian plane in presence of third order coma of amounts $\Phi_{max} = 0{\cdot}3\lambda$, 1λ, $2{\cdot}4\lambda$, 5λ, 10λ,. *After* Nienhuis, *reprinted from* Zernike (1948).

Figure 17. Images in central plane in presence of third order astigmatism of amounts $\Phi_{max} = 1{\cdot}4\lambda$ $2{\cdot}7\lambda$, $3{\cdot}5\lambda$, $6{\cdot}5\lambda$. *After* Nienhuis, *reprinted from* Zernike (1948).

of relations connecting the various transmission factors and the intensity distribution in the image. One of these relations obtained by a formal application of Plancherel's theorem connects P with the moments of the energy distribution of the image. Duffieux (1948) stated that these moments when taken about the centre of gravity of the distribution are the same for the geometrical and the physical image. As has been pointed out by E. H. Linfoot*, Duffieux' analysis is rendered invalid by failure to take into account the divergence of certain

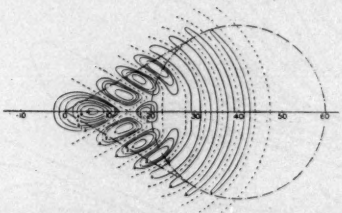


Figure 20. Isophotes in Gaussian plane in presence of third order coma, of amount $\Phi_{max}=3{\cdot}2\lambda$. *After* Kingslake (1948).

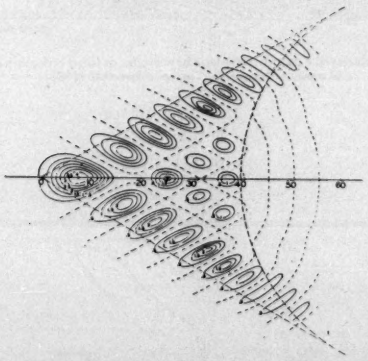


Figure 21. Isophotes in Gaussian plane in presence of third order coma, of amount $\Phi_{max}=6{\cdot}4\lambda$ *After* Kingslake (1948).

integrals; that the result also is incorrect can easily be seen by considering the moments of an error-free image. Several other conclusions of Duffieux are based on faulty mathematical arguments and appear to be incorrect.

Kingslake (1948) investigated the diffraction structure of third order comatic images by numerical integrations. He demonstrated that the image is of the size and shape determined by ordinary geometrical optics and that the main effect of diffraction is to break up the image into an elaborate fine structure of dots and lines of light. We give here two of his valuable diagrams (Figures 20 and 21) showing the isophotes for coma of fairly large amounts.

* *Math. Rev.*, 1946, **7**, 269; *Ibid.*, 1949, **10**, 220.

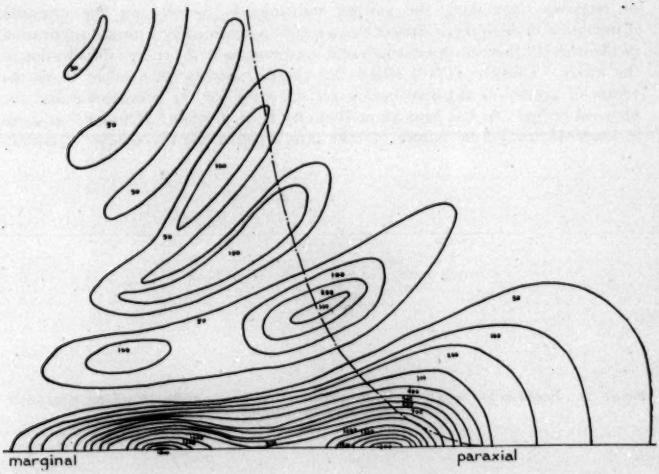


Figure 22. Isophotes in a meridional plane in presence of third order spherical aberration, of amount $\Phi_{max}=4\lambda$. *After* Maréchal (1951).

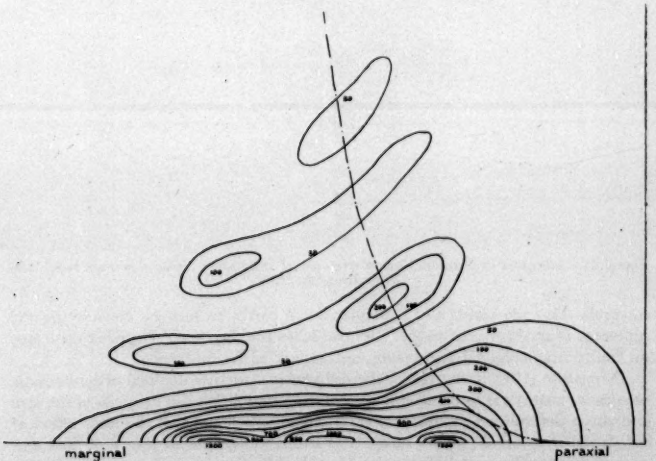


Figure 23. Isophotes in a meridional plane in presence of third order spherical aberration, of amount $\Phi_{max}=6\lambda$. *After* Maréchal (1951).

Maréchal (1947b, 1948) designed and constructed a machine for rapid evaluation of double integrals of the type occurring in the investigations of aberration diffraction effects. To determine the intensity at each point of the image, operations taking about 8–10 minutes are needed. The precision of the apparatus is quite adequate for practical purposes, the estimated inaccuracies ΔI being as follows ($I=1$ at the centre of the Airy pattern): $\Delta I<0{\cdot}01$ when $I\sim 1$, $\Delta I<0{\cdot}003$ when $I\sim 0{\cdot}1$, $\Delta I<0{\cdot}001$ when $I\sim 0{\cdot}01$. With the help of this

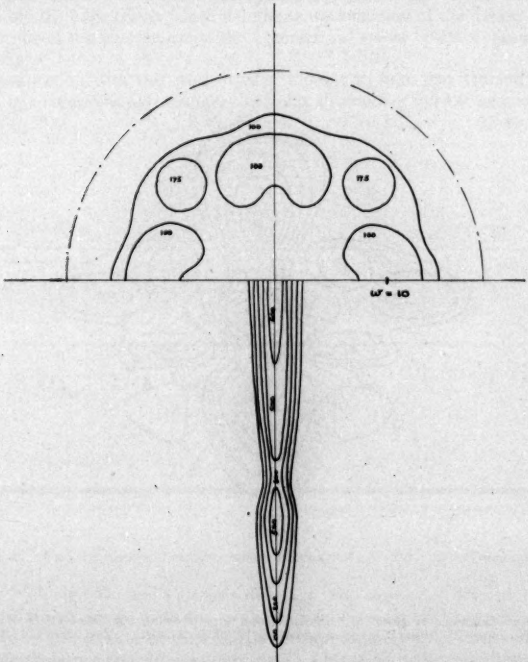


Figure 24. Isophotes in presence of third order astigmatism, of amount $\Phi_{max}=1{\cdot}5\lambda$ in central plane (upper half), in plane containing a focal line (lower half). *After* Maréchal (1951).

integrator Maréchal (1948, 1951) obtained a number of important isophote diagrams some of which are shown in Figures 22–25.* Most of these diagrams show the effects of aberrations in the range 1λ to 10λ for which available series expansions are not well adapted. The calculations performed with the aid of this integrator were also compared by Maréchal with photometric measurements of the intensity distribution in actual images; good agreement was found.

* I am greatly indebted to Dr. Maréchal for giving me access to these figures before the publication of his own paper.

It is well known that in the approximations of Kirchhoff's scalar theory, the disturbance produced by the passage of spherical waves through an aperture opening may be regarded as due to the combined effect of ' geometrical waves ' and ' diffraction waves ' emerging from the aperture edge.* Nienhuis (1948) assumed that this result holds (and also that the ' diffraction waves ' possess analogous properties) when the incident waves are not strictly spherical† and

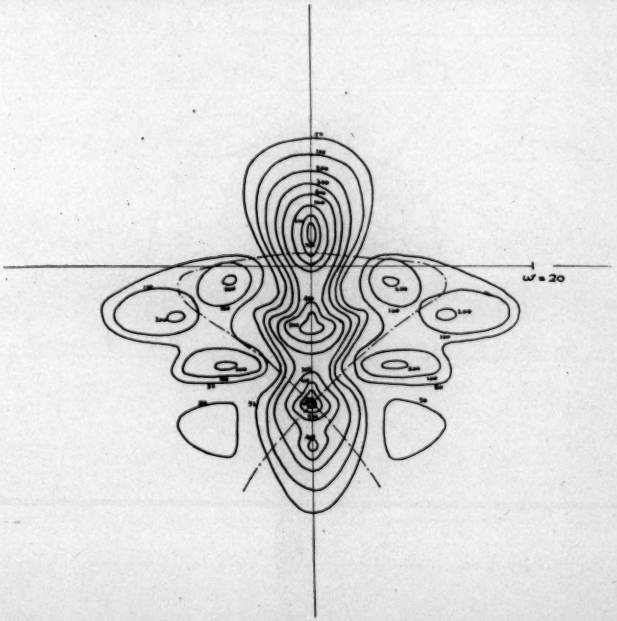


Figure 25. Isophotes in plane of astigmatic focal line, showing combined effect of spherical aberration (2λ), coma (1λ) and astigmatism (1λ) of third order. *After* Maréchal (1951).

applied it to discuss some general features of diffraction images in the presence of third order aberrations.

In the case of astigmatism, Nienhuis showed by a geometrical argument that the ' diffraction waves ' from a circular edge give rise to an asteroid pattern. This pattern is partly obstructed by the effect of the ' geometrical waves '. Its

* Ramachandran (1945) showed that the integral expressing the edge effect can be derived in a very simple manner if the variation of the amplitude of the secondary wavelets with direction is neglected. This idea was also used by Kathawate (1945) to discuss the effect of diffraction of spherical waves at boundaries and obstacles of various forms.

† Many assertions in Nienhuis' thesis and in other literature give the impression that the transformation of the diffraction integral into an integral expressing the edge effect and a term expressing the geometrical effect has been carried out in the general case, but an examination of the relevant papers reveals that the published proofs relate to spherical and plane waves only.

shape is in the first approximation independent of the focal setting and is therefore more apparent in a receiving plane through one of the focal lines (see Figure 12).* In the presence of coma, rays from opposite points of each aperture zone meet the Gaussian plane at the same point. Nienhuis found that the fringe structure of the comatic pattern can be ascribed to the interference of these rays, in agreement with his observation that the fringes disappear when half of the light is blocked out by using a semicircular aperture (Figure 10). For spherical aberration Nienhuis found that in the receiving plane through the paraxial focus it is mainly the edge effects which determine the structure of the pattern whilst in other planes the interference of the ' geometrical waves ' plays a considerable part.

In connection with this work it is of interest to note that Durand (1949) studied in a comprehensive manner the fringe structure of third order aberration

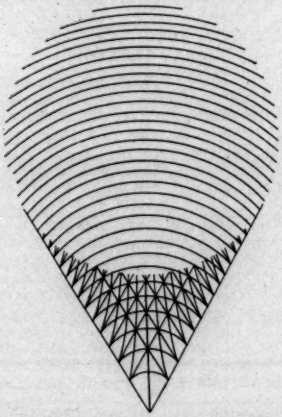


Figure 26. Fringe structure of comatic pattern of amount $\Phi_{max}=8{\cdot}3$. *After* Durand (1949).

figures. Though Durand's analysis does not take account of diffraction it has many points of contact with the work of Nienhuis. Durand's results may be expected to give good approximations to the appearance of the image whenever edge effects do not play an important part. This is particularly well illustrated by the case of coma (Figure 26).

Françon (1948 a, b) discussed the effects on the image due to a certain type of non-homogeneity in the glass of an objective. He found that in the presence of filaments or veins whose refractive index differs slightly from the bulk of the glass, the effects are negligible if the path differences they introduce are less than $\lambda/8$ for $p=50$, $\lambda/6$ for $p=100$ and $\lambda/4$ for $p=200$, p being the ratio of the

* Some of Nienhuis' photographs of diffraction images are also given by Zernike (1948). The corresponding values of the aberration constants differ in several cases and some appear to be incorrect. In particular it appears that the aberration constants for the first two comatic patterns of Figure 16 should be larger. Values of Φ_{max} based on Nienhuis' data are given in Figures 16 and 17.

diameter of the objective to the width of the filament. When one observes two neighbouring objects of widely differing brightness the tolerances are much more severe.

Epstein (1949) took up once more the problem treated by Lommel (1885) (see § 3.1) and found methods for the numerical evaluation of the associated diffraction integral, which possess some advantages over those of Conrady and Buxton.

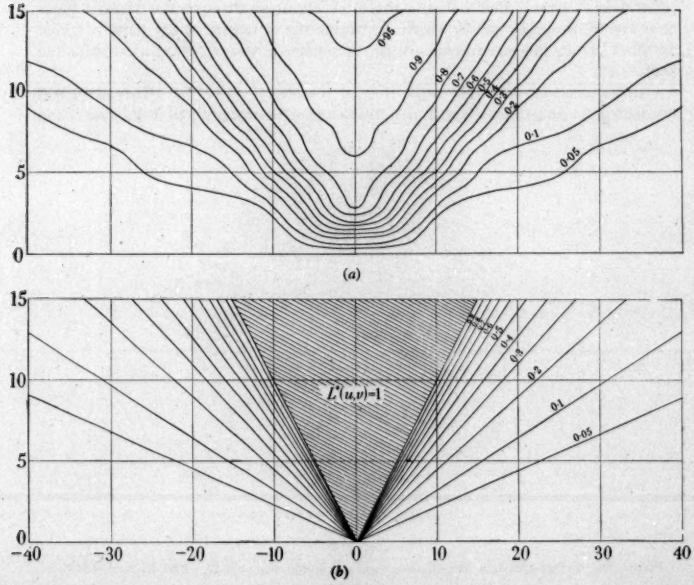


Figure 27.

$$u=\frac{2\pi}{\lambda}\left(\frac{a}{f}\right)^2\Delta f, \quad v=\frac{2\pi}{\lambda}\frac{a}{f}r,$$

a = radius of exit pupil, f = focal length, Δf = amount of defocusing.

Contour lines for the fraction of the total illumination inside circles of radius r, centred on axis, in receiving planes near focus of spherical waves issuing from a circular aperture (*a*) according to diffraction theory, (*b*) according to geometrical optics. *After* Wolf (1950).

The distribution of the total illumination in out-of-focus images formed by spherical waves issuing from a circular aperture was investigated by Wolf (1951). He derived expressions for the fraction of the total illumination present inside circles of given radii in the receiving plane and applied them to study the distribution in a few selected planes near the focus. Wolf also gave a diagram (see Figure 27) showing the corresponding three-dimensional distribution and compared it with the prediction of geometrical optics. His results, together with those of Lommel, Zernike and Nijboer, provide a mathematical basis for a

quantitative discussion of the imaging properties in optical systems where, as for example in a well corrected refracting telescope objective, the chromatic variation of focus is the only appreciable aberration.

Mme. Regnier (1950) studied the influence of spherical aberration on the limiting visibility of Foucault test objects. She used a relation due to Duffieux and Lansraux between the pupil function $P(p, q)$, the transmission factor $d(p)$ and the limiting visibility, and calculated $d(p)$ for spherical aberration of amounts $\lambda/4$, $\lambda/2$, $3\lambda/4$, λ and for different forms of exit pupil. The results yield conclusions about the dependence of the limiting visibility of test objects on the aberration of the objective.

ACKNOWLEDGMENTS

The author gratefully acknowledges the permission granted by various authors and publishing bodies for the reproduction of diagrams and photographs.

REFERENCES

AIRY, G. B., 1835, *Trans. Camb. Phil. Soc.*, **5**, 283.

BEREK, M., 1926, *Z. Phys.*, **40**, 421.

BORN, M., 1932, *Naturwissenschaften*, **20**, 921; 1933, *Optik* (Berlin: Springer), p. 202.

BOUGHON, P., DOSSIER, B., JACQUINOT, P., 1946, *C. R. Acad. Sci., Paris*, **223**, 661.

BUXTON, A., 1921, *Mon. Not. R. Astr. Soc.*, **81**, 547; 1923, *Ibid.*, **83**, 475.

CONRADY, A. E., 1919, *Mon. Not. R. Astr. Soc.*, **79**, 575.

VAN DER CORPUT, J. G., 1948, *Proc. K. Akad. wet. Amst.*, **51**, 6.

COUDER, A., 1944, *C. R. Acad. Sci., Paris*, **219**, 677.

DUFFIEUX, P. M., 1945 a, *C. R. Acad. Sci., Paris*, **220**, 846; 1945 b, *Ibid.*, **220**, 911; 1946 a, *L'Intégrale de Fourier et ses Applications a l'Optique* (Besançon: Faculté des Sciences); 1946 b, *C. R. Acad. Sci., Paris*, **222**, 1482; 1947, *Ann. Phys., Paris*, **2**, 95; 1948, *C. R. Acad. Sci., Paris*, **226**, 1257.

DUFFIEUX, P. M., and LANSRAUX, G., 1945, *Rev. Opt.* (*théor. instrum.*), **24**, 65, 151, 215.

DURAND, E., 1949, Contribution in *La Théorie des Images Optiques* (Paris: Éditions de la *Revue d'Optique*), p. 127.

EPSTEIN, L., 1949, *J. Opt. Soc. Amer.*, **39**, 226.

FRANÇON, M., 1944, *Cah. Phys.* No. 26, 15; 1947 a, *Rev. Opt.* (*théor. instrum.*), **26**, 354; 1947 b, *Ibid.*, **26**, 369; 1948 a, *Ibid.*, **27**, 157; 1948 b, *Ibid.*, **27**, 595, 761; 1949, *C. R. Acad. Sci., Paris*, **229**, 293.

HOPKINS, H. H., 1943, *Proc. Phys. Soc.*, **55**, 116; 1944, *Ibid.*, **56**, 48; 1949, *Proc. Phys. Soc.* B, **62**, 22.

JACQUINOT, P., BOUGHON, P., DOSSIER, B., 1949, Contribution in *La Théorie des Images Optiques* (Paris: Éditions de la *Revue d'Optique*), p. 183.

VAN KAMPEN, N. G., 1949, *Physica*, **14**, 575.

KATHAWATE, Y. V., 1945, *Proc. Indian. Acad. Sci.* A, **21**, 177.

KINGSLAKE, R., 1948, *Proc. Phys. Soc.*, **61**, 147.

KÖNIG, A., 1929, *Handb. Exp. Phys.*, **20**, Part II, 141.

KORRINGA, J., NIJBOER, B. R. A., and MARÉCHAL, A, Document S.O. 48-8, Commission Internationale d'Optique (Paris: Institute d'Optique).

LANSRAUX, G., 1946, *C. R. Acad. Sci., Paris*, **222**, 1434; 1947 a, *Rev. Opt.* (*théor. instrum.*), **26**, 24; 1947 b, *Ibid.*, **26**, 278; 1949, Contribution in *La Théorie des Images Optique* (Paris: Éditions de la *Revue d'Optique*), p. 194.

VON LAUE, M. V. 1928, *Handb. Exp. Phys.*, **18**, 211.

LOMMEL, E., 1885, *Abh. Bayer. Akad.*, **15**, Abth. 2, 229; 1886, *Ibid.*, **15**, Abth. 3, 529. Shortened account given in GRAY, A., MATHEWS, G. B., and MACROBERT, T. M., 1931, *A Treatise on Bessel Functions*, 2nd Edit. (Macmillan: London).

LUNEBERG, R. K., 1944, *Mathematical Theory of Optics* (Providence, R.I.: Brown University).

MARÉCHAL, A., 1944 a, *C. R. Acad. Sci., Paris*, **218,** 345; 1944 b, *Ibid.*, **218,** 395; 1944 c, *Cah. Phys.*, No. 26, 1; 1947 a, *Rev. Opt. (théor. instrum.)*, **26,** 257; 1947 b, *J. Opt. Soc. Amer.*, **37,** 403; 1948, *Rev. Opt. (théor instrum.)*, **27,** 73; 1951, *Ibid.*, in the press.

MARTIN, L. C., 1922, *Mon. Not. R. Astr. Soc.*, **82,** 310; 1946, *Photogr. J.*, **86 B,** 47.

MICHELSON, A. A., 1905, *Phil. Mag.*, **9,** 506.

NIENHUIS, K., 1948, *Thesis*, University of Groningen.

NIENHUIS, K., and NIJBOER, B. R. A., 1949, *Physica*, **14,** 590.

NIJBOER, B. R. A., 1942, *Thesis*, University of Groningen; 1943, *Physica*, **10,** 679; 1947, *Ibid.*, **13,** 605.

OSTERBERG, H., and WILKINS, J. E., 1949, *J. Opt. Soc. Amer.*, **39,** 553.

OSTERBERG, H., and WISSLER, F. C., 1949, *J. Opt. Soc. Amer.* **39,** 558.

PICHT, J., 1925, *Ann. Phys., Lpz.*, **77,** 685; 1926, *Ann. Phys., Lpz.*, **80,** 491. (Also 1931, *Optische Abbildung* (Braunschweig : Vieweg).)

RAMACHANDRAN, G. N., 1945, *Proc. Indian Acad. Sci.* A, **21,** 165.

RAYLEIGH, 1879, *Phil. Mag.*, **8,** 403. (Also *Scientific Papers*, **1,** 428.)

RÉGNIER SIMONE M. J., 1950, *Rev. Opt. (théor. instrum.)*, **29,** 315.

RICHTER, R., 1925, *Z. Instrumkde*, **45,** 1.

SCHWARZSCHILD, K., 1897, *S.B. München Akad. Wiss.* (*Math.-Phys. Kl.*); **27,** 271; 1905, *Astro. Mitt. Köngl. Sternwarte, Göttingen*, **9.** (Also *Abh. K. Gess. Wiss., Göttingen, Math.-Phys. Kl.*, **4.**)

SLEVOGT, H., 1949, *Optik*, **4,** 349.

STREHL, K., 1894, *Theorie des Fernrohrs* (Leipzig: Barth).

STEWARD, G. C., 1925, *Phil. Trans. Roy. Soc.* A, **225,** 131; 1926, *Trans. Camb. Phil. Soc.*, **23,** 235. (Also 1928, *The Symmetrical Optical System*, Cambridge Tracts, No. 25.)

STRAUBEL, R., 1902, *Phys. Z.*, **4,** 16, 74; 1931, *8th International Congress on Photography*, Dresden (Leipzig: Barth), p. 353.

STRUVE, H., 1886, *Mém. Akad. Sci., St. Pétersbourg* (7), **34,** No. 5, 1.

TORALDO DI FRANCIA, G., 1946, *Atti Fond.* " *Giorgio Ronchi* ", **1,** 136.

VÄISÄLÄ, Y., 1922, *Annales Univ. Fennicae Aboensis*, A, **1,** 108.

WANG TA-HANG, 1941, *Proc. Phys. Soc.*, **53,** 157.

WILKINS, J. E., 1950, *J. Opt. Soc. Amer.*, **40,** 222.

WOLF, E., 1951, *Proc. Roy. Soc.* A, **204,** 533.

ZERNIKE, F., 1934, *Physica*, **1,** 689 ; 1948, *Proc. Phys. Soc.*, **61,** 158.

ZERNIKE, F., and NIJBOER, B. R. A., 1949, Contribution in *La Théorie des Images Optiques* (Paris : Éditions de la *Revue d'Optique*), p. 227.

THE SPECTRUM OF THE AIRGLOW AND THE AURORA

By A. B. MEINEL*
Yerkes Observatory, University of Chicago

CONTENTS

	PAGE
§ 1. Introduction	121
§ 2. The spectrum of the night glow	122
2.1. Ultra-violet region from 3000 A. to 4000 A.	123
2.2. Visible region from 4000 A. to 7000 A.	124
2.3. Near infra-red region from 7000 A. to 11,000 A.	126
2.4. Infra-red region beyond 11,000 A.	129
2.5. Continuum	130
§ 3. The spectrum of the twilight glow	130
§ 4. The spectrum of the aurora	133
4.1. Ultra-violet region from 3000 A. to 4000 A.	133
4.2. Visible region from 4000 A. to 7000 A.	134
4.3. Infra-red region from 7000 A. to 11,000 A.	135
4.4. Microwave region	136
4.5. Atomic emission features	137
§ 5. Heights of the night-glow emissions	140
§ 6. Temperature determinations from the night-glow emissions	143
§ 7. Temperature determinations from the auroral emissions	144
§ 8. Conclusions	144
Acknowledgments	145
References	145

ABSTRACT. The spectrum of the night glow has been investigated from 3000 A. to 14 μ. While the majority of the emissions have been identified, numerous unidentified bands remain. The molecules O_2, N_2 and OH contribute emissions between 3000 A. and 4·5 μ.

The enhancement of the N_2^+ bands and the [O I] and Na D-lines during twilight is generally attributed to resonance scattering. Several observers have suggested that molecular dissociation can explain certain of these enhancements.

The major features of the auroral spectrum are well identified. There is still uncertainty as to the existence of ionized atomic emissions. Detailed observation of hydrogen emissions has clearly established that protons of extra-terrestrial origin are the primary particles of the aurora.

Much progress has been made in height and temperature determinations for the airglow. Outstanding discrepancies, however, still exist for the Na D-lines and for the Herzberg and Vegard–Kaplan bands.

§ 1. INTRODUCTION

THE fact that the background of the night sky is not entirely dark has been known for centuries, while references to aurorae extend back into the dim recesses of Nordic antiquity. Yet it has only been since the turn of the twentieth century that these phenomena have been the subject of systematic study. The presence of an emission spectrum for the aurora was recognized as early as 1867, when Ångström observed and measured the wavelength of the green line. The presence of an emission spectrum from the night sky was noted in 1895, purely as a nuisance, since it interfered with astronomical spectroscopic

* Formerly at the Lick Observatory, University of California.

observations. W. W. Campbell (1895) discovered visually the presence of the green line in the normal night sky, many years before it was re-discovered photographically by V. M. Slipher. He writes that "visual observations of faint nebular and comet spectra are often interfered with by the aurora and sky spectra. . . . I have looked for it (green line) in all parts of the sky . . . always with success". The light of the night sky as a detached phenomenon of physical interest was recognized by the pioneering investigations of Fabry (1910), V. M. Slipher (1919), Lord Rayleigh (1920) and Dufay (1923). For a historical record of night sky and auroral research prior to the developments discussed in this paper, the reader is referred to articles by Dufay (1928), Fabry, Dufay and Cojan (1934), Déjardin (1935), Elvey (1942), Dufay (1943), and Swings (1949).

Progress in the investigation of the light of the night sky and aurora has greatly accelerated since the end of World War II. Instrumental advances have extended the range of observation on the short wavelength side to the atmospheric limit in the ultra-violet, and on the long wavelength side into the microwave region. As a consequence, new problems have arisen faster than old problems were solved. Even as this report goes to press, ultra-fast spectrographs of unprecedented proportions and resolution are under construction in several countries. Observations with these instruments should do much toward bringing about a definitive solution of the outstanding auroral and night sky problems within the next few years.

A study of the light of the night sky and aurora should yield important information concerning the physics of the upper atmosphere. The realization of this goal has been hampered because it was not possible to ascribe a height at which each phenomenon originates with any reasonable degree of certainty. Recent technical and theoretical advances, however, give promise of the removal of this barrier. Once it is possible to assign a height to each emission, the temperature and the molecular and atomic processes can be integrated into a unified physical picture of the upper atmosphere, coordinated with the extensive data obtained by rocket flights and by ionospheric and meteor observations.

A brief note on nomenclature is appropriate at this time. Ever since the study of the diffuse light of the night sky began, authors have been handicapped by the long appelation 'light of the night sky' when quite often only the atmospheric emission component was implied. Recently the term 'airglow' has been introduced by Elvey (1950). This term, however, does not distinguish between the several distinct types of airglow. The author has therefore added the specific terms 'night glow' and 'twilight glow'. Current work with rockets may, in fact, make the term 'day glow' as common as 'night glow'.*

§2. THE SPECTRUM OF THE NIGHT GLOW

The photography of the spectrum of the night glow has always been difficult because of the low surface brightness of the night sky, whose light is so faint that long exposures and low dispersions must be used. In the red and infra-red regions, the investigator is further handicapped by the extremely low dispersion of prismatic spectrographs and by the slowness of infra-red sensitive photographic emulsions. In recent years, however, observations of these regions have been made with grating spectrographs and very large multiple prism spectrographs.

* *Note added in proof.* Professor Sydney Chapman has recently proposed a similar nomenclature in the October 1940 issue of the *Bulletin of the American Meteorological Society*.

The identification of the airglow emissions is made difficult because of the uncertainty of wavelengths and emission profiles due to the low dispersions. These difficulties are multiplied by the great number of observed bands, especially in the blue and ultra-violet, and the resultant blending.

2.1. *Ultra-violet Region from* 3000 A. *to* 4000 A.

The ultra-violet region of the spectrum of the night glow is a region rich in emission features, the majority of which stand out clearly against the background continuum. Observation of the ultra-violet region was not possible until Dufay (1934) employed a spectrograph equipped with quartz optics, since flint prism spectrographs seldom could attain wavelengths shorter than 3800 A. Recent spectra of the night glow by Barbier (1945) using a two-prism F/2 quartz spectrograph have provided excellent information on the ultra-violet regions. A critical summary of all wavelength measures in this region is given by Dufay and Déjardin (1946).

Herzberg Bands of O_2.

Dufay (1941) and Swings (1943) noticed that the majority of the bands in the region of 3000 A. to 4000 A. could be identified with the forbidden $^3\Sigma_u^+ - ^3\Sigma_g^-$ transition of O_2, discovered in absorption studies by Herzberg (1932). The identification of these bands with the Herzberg bands is especially attractive since the Herzberg bands have a low excitation potential (4·5 ev.) and represent a forbidden transition to the ground state of the molecule. This identification has been discussed in detail by Dufay (1946), Barbier (1945, 1947 a) and Déjardin (1948).

The bands associated with this system group closely around the expected Franck–Condon parabola. The bands appear to be due to the progressions $v' = 1$ to 4, with a few greater than 4. It is sometimes difficult to attribute a definite (v', v'') to the bands, since the wavelengths of the night-glow emissions are not known with sufficient accuracy. As a consequence, the same emission feature is often ascribed to several (v', v'') values. Since these emissions may, in fact, be blends of several transitions, it is imperative that higher dispersion spectrographs be used to facilitate a definite answer. The question of blends is also of the greatest importance when emission profiles are used to determine temperatures.

Vegard–Kaplan Bands of N_2.

A moderate number of weak emissions in the ultra-violet region of the night glow have been identified with members of the Vegard–Kaplan bands. The strongest bands of this system occur, however, in the visible region, and will be discussed later.

Barbier Bands of (*CO*?).

Aside from the Herzberg and Vegard–Kaplan bands, a considerable number of bands in the ultra-violet and blue regions remain unidentified. Barbier (1947 a) developed a novel method of analysis, by means of which he found that the unidentified bands could be grouped in a system having vibrational constants close to those of CO. About 25 night-glow emissions can be identified in this manner. Swings (1949), however, has raised objections to this identification on the basis that : (i) although CO is a minor constituent of the atmosphere, the

CO bands are almost as intense as those from O_2 or N_2; (ii) the value of $\omega_e = 2560\ cm^{-1}$ appears high for CO. The first objection can now be held in abeyance by referring to the infra-red OH bands, which originate from a minor constituent, but which produce the most intense emission features in the spectrum of the night glow. The determining factor appears to be whether or not an efficient two-body excitation mechanism can exist. The second objection by Swings, however, is alone sufficient to question the reality of the proposed CO identification.

Other Bands.

The presence of the Schumann–Runge bands, suggested by Kaplan (1935) is doubtful, although some coincidences can be found. These coincidences appear to be fortuitous, since these emissions can be attributed to the band system already discussed. Similarly, the large number of coincidences found by Déjardin and Bernard (1938) between night-glow emissions and OH electronic bands appear to be fortuitous.

2.2. *Visible Region from* 4000 A. *to* 7000 A.

The visible region of the spectrum of the night glow is characterized by the presence of atomic lines in the green, yellow and red, and of numerous broad bands superimposed on a strong continuum in the blue. Because of its accessibility to ordinary spectrographs the visible region has been studied extensively. Although numerous advances have been made in the interpretation of the spectrum of this region of the night glow, outstanding problems remain for future work.

Atomic Lines.

The conclusive identification of atomic lines in the night-glow spectrum was a long and difficult task. While the presence of the green line of [OI] at 5577 A. was recognized thanks to the accurate measurement by Babcock (1923) and the subsequent laboratory identification by McLennan and Shrum (1924), the presence of the red [OI] lines and the Na D-lines was long debated. For a considerable time instrumental limitations prevented the resolution of the red [OI] doublet at 6300·2 A. and 6363·9 A. The final proof of the existence of the Na D-lines was not made until the announcement by Bernard (1938 a) of the results of an interferometric study. Subsequent work by Cabannes, Dufay and Gauzit (1938) confirmed this result.

The atomic lines that are certainly present are the [OI] ^1D–^1S transition at 5577 A., the [OI] ^3P–^1D transition at 6300·2 A. and 6363·9 A., and the Na D-lines. The [NI] lines at 3466·5 A., ^4S–^2P, and 5197·8 A. and 5200·1 A., ^4S–^2D, are definitely absent. Occasional identification of the 3303 A. Na doublet and other atomic metallic lines, such as CaI, CaII, AlI, etc., may be attributed to low dispersion wavelength coincidences with molecular bands. Certainly, there are no sharp emission lines in the blue as there are in the green and red.

On low dispersion spectra, the Na D-line often appears to be diffuse (Vegard and Tonsberg 1935), especially when the Na D-line is weak. A similar phenomenon occurs for the red [OI] lines. The cause of this effect, which was responsible for some of the uncertainty in the Na D identification, was found by Meinel (1950 c, d) and Cabannes, Dufay and Dufay (1950). The identification of the OH bands in the infra-red permitted the prediction of the positions of other bands. Meinel (1950 c, d) computed and observed that the (8, 2) OH band

origin was at 5890 A. On low dispersion spectra the R and Q branches of this band are hidden by the Na D-lines. The P branch, however, is observable at 5965 A. and accounts for the wing on the D-lines. Spectra recently obtained by Cabannes, Dufay and Dufay (1950) independently confirmed this interpretation, since they were able to resolve the R branch from the D-lines and also resolve the rotational structure of the P branch. The (9, 3) OH band at 6258 A. causes the similar effect with the red [OI] lines.

The absence of the nitrogen forbidden lines has posed many questions. In accounting for the presence or absence of atomic lines of OI and NI, it is necessary to consider the abundances of the atomic constituents, the energy sources, and the transition probabilities. In the case of oxygen, the vertical distribution has been thoroughly investigated, most recently by Penndorf (1949). Molecular oxygen is readily dissociated by solar radiation, to such an extent that all oxygen above 100 km. is in the atomic form. During the night this oxygen remains largely in atomic form above 120 km., the largest amount of recombination occurring between 90 and 120 km. Atomic oxygen will therefore be very abundant during the night for any excitation mechanism in the upper atmosphere.

The question as to the state of nitrogen in the upper atmosphere is still uncertain. Photodissociation is much more difficult to achieve for N_2 than for O_2. Radiations shorter than 750 A. are required to photodissociate N_2, while radiation at 2000 A. can dissociate O_2. Herzberg (1949) has suggested that radiation of $\lambda < 1400$ A. could lead to a predissociation of N_2 by a forbidden transition from the ground state to a high vibrational level of the $A^3\Sigma$ state. The absorption in such a transition, however, is apt to be very weak. The difficulty of dissociating molecular nitrogen, the absence of [NI] emissions in the night glow, and the presence of N_2^+ emission very high in sunlit aurorae have therefore been the chief reasons for assuming that nitrogen in the upper atmosphere is mainly in the molecular state.

The absence of the [NI] lines, however, is not definite. For an emission to be present in the night glow, an efficient excitation mechanism must be present. Furthermore, the requirement of an energy source may limit each region of emission to a relatively thin stratum. Regardless of the height of this stratum, the lifetimes of the excited states are important. For [OI], the lifetimes are 0·44 second for the ^1S level, and 96 seconds for the ^1D level. For nitrogen the lifetimes are 2·1 seconds for the ^2P level and 8 to 12 hours for the ^2D level. The long lifetime of the NI ^2D level would effectively prohibit the emission of the 5200 A. [NI] doublet except at extreme heights. Under favourable excitation conditions, however, the 3466 A. [NI] line could be radiated from regions as low as 100 km. Since the excitation potential required for the 3466 A. [NI] line is lower than for the 5577 A. [OI] line, atomic nitrogen must be very rare below 150 km.

Bates (1950) has shown that the most probable excitation mechanism for the Vegard–Kaplan bands requires the presence of atomic nitrogen. Consequently, the firmer establishment of the identification of the Vegard–Kaplan bands and a reliable height determination would materially assist in a solution of the problem of the state of nitrogen.

The presence of H_α emission in the spectrum of the night glow has been suggested by Elvey (1950). The broad emission centred at 6563 A. observed by Elvey, but with very low dispersion, was interpreted as a Doppler-broadened H_α emission, similar to the hydrogen emission lines observed by Gartlein (1950) in some aurorae. Meinel (1950 c) has shown that the emission at 6560 A., observed

by Babcock (1939), is due to the (6, 2) OH band. The centre-of-gravity of the P branch of this band is at 6563·2 A., and under low dispersion could be identified as H_α. The interpretation by Meinel (1950 c) has been confirmed by recent observations by Gartlein (private communication). The spectrograph employed by Gartlein for current auroral studies easily resolves the P branch lines of the (6, 2) OH band. H_α would fill the region between P(7/2) and P(9/2). He has not observed an emission of this position with as much as half the intensity of the adjacent OH rotational lines.

First Positive Bands of N_2.

Apart from the atomic lines, numerous weaker emissions are observed in the region from 5000 A. to 7000 A. These emissions were ascribed by numerous authors to the first Positive bands of N_2. These identifications indicated some selectivity at the $v'=6$ and $v'=7$ levels. Recent work by Meinel (1950 c, d) and Cabannes, Dufay and Dufay (1950) has shown that all of the bands between 5000 A. and 7000 A. are accounted for, even in detailed structure, by the $\Delta v=5$ and $\Delta v=6$ OH band sequences. These investigations have therefore conclusively eliminated the first Positive bands from the spectrum of the night glow.

Vegard–Kaplan Bands of N_2.

The strongest bands in the blue and violet region have been identified as the Vegard–Kaplan $A^3\Sigma$–$X^1\Sigma$ intercombination bands. This system accounts for emissions from 5000 A. to 3100 A.; however, the part beyond 3800 A. can account for only very weak emissions. The strongest bands from 4535 A. to 3949 A. belong to the sequences $\Delta v=-11$ and $\Delta v=-12$, and follow the expected Franck–Condon intensity rule. The bands show some selectivity around $v'=2$, 3 and 4 levels; however, there are also some originating from $v'=1$, 2 and 5 to 9 levels. Only the strongest bands identified with the Vegard–Kaplan system show a degradation towards the red. The profiles of the weaker bands may be affected by blends.

2.3. *Near Infra-red Region from* 7000 A. *to* 11,000 A.

The strongest emissions of the night glow were only recently discovered. Stebbins, Whitford and Swings (1944) found an intense radiation having an effective wavelength of 10,440 A. by means of a photoelectric cell and suitable infra-red filters. The presence of a radiation near 10,300 A. was reported by Hermann, Hermann and Gauzit (1942); however, the fact that they have not been able to confirm their single observation, even under the most favourable conditions, indicates that their spectroscopic observation must have been in error. Dufay (1949 a) points out that Hermann, Hermann and Gauzit did not observe the strong features in the region of 8000 A. later observed by Meinel (1948) and Dufay (1949 a), although their 'Agfa 800' plate was more sensitive by a large factor at that wavelength than at 10,000 A. In fact, a hypersensitized Eastman 1-Z emulsion and and F/1 spectrograph having a dispersion of 2000 A/mm. would require about 1,000 hours to record the emission at 10,000 A.

First Positive Bands of N_2.

The emission in the 10,440 A. region was attributed by Stebbins, Whitford and Swings (1944, 1945) to the (0, 0) band of the first Positive system of N_2. The fact that the heat of dissociation of N_2 of 7·38 ev. is higher than the excitation energy of the $v'=0$ level of the upper state of the first Positive bands ($^3\Pi$), but

lower than the $v'=1$ level provided a possible mechanism to account for a strong (0, 0) transition. This identification of the 10,440 A. emission, while very satisfactory from the standpoint of the excitation mechanism, raised difficulties when the intensity distribution of the Vegard–Kaplan bands was considered. Since a strong (0, 0) first Positive band would heavily populate the $v'=0$ upper level ($^3\Sigma$) of the Vegard–Kaplan bands, the $(0, n)$ Vegard–Kaplan bands should be the strongest according to calculations by Bates (1949 a). Observationally, the Vegard–Kaplan bands originating from $v'=2$ and $v'=3$ are the strongest.

The forbidden ^{2}D–^{2}P transition of NI (10,398–10,407 A.) could not contribute significantly to the emission at 10,440 A., since the other [NI] transitions are absent.

Meinel Bands of OH.

The discovery by Meinel (1950 a) that the molecule OH gives rise to the prominent features in the photographic infra-red has offered a satisfactory solution to the problem presented by the 10,440 A. emission. A complete analysis by Meinel (1950 d, e), in co-operation with Herzberg, has shown conclusively that the emissions from 7000 A. to 9000 A. are due to the rotation–vibration spectrum of the OH molecule. This identification marks the first instance of the observation of a vibrational emission spectrum from a free radical. The physical situation is remarkable since these intense bands originate from a minor constituent of the upper atmosphere. The analysis was confirmed by the prediction of other OH bands in the visible region of the night glow, found subsequently by Meinel (1950 c) and Cabannes, Dufay and Dufay (1950).

With the identification of the OH bands it was possible to predict the position of other sequences of this system. The $\Delta v=4$, 3, 2 and 1 sequences extend into the photographically unobserved infra-red region, as far as 4·5 μ. It is certain that these bands are much more intense than those currently observed in the photographic region. The 10,440 A. radiation is composed of an overlap by the $\Delta v=3$ and $\Delta v=4$ sequences. In all, eleven OH bands are included in the spectral region used by Stebbins, Whitford and Swings.

It should be noted that this identification of the 10,440 A. emission does not exclude the presence of the (0, 0) band of the first Positive system of N_2. It does, however, exclude the association of a large proportion of the energy observed at 10,440 A. to the (0, 0) band. Spectroscopic observations of considerable resolution will be required to determine whether or not the (0, 0) band is present. Some preliminary observations, with the aid of a photoelectric detector and a fast spectrograph, have been made by Meinel and Smith (unpublished) and Kron and Smith (private communication); however, the resolution was insufficient to permit the detection of any emission weaker than the nearby OH bands.*

The question of an excitation mechanism inducing a minor constituent of the atmosphere to emit the most intense radiations has aroused considerable theoretical

* *Note added in proof.* Krassovsky (1949, 1950) obtained spectra of the infra-red between 8000 A. and 11,000 A. using a prismatic spectrograph and a Cs–O–Ag electron image converter. Kron (1950) has made a study of this region with a quartz monochromator and CsO photocell. These two investigators are in complete agreement as to the emissions. The wavelengths, however, differ slightly as a consequence of the feeble dispersion employed. Dufay (1950) and Kron (1950) have found that the emissions are due to the Meinel bands of OH. Dufay (1950) suggests that the emission reported by Krassovsky at 9976 A. could be due to the (0, 2) O_2 band, except that the intensity is higher than can be explained in view of the intensity of the (0, 1) band. It is probable that the 9976 A. emission is a blend of the (0, 2) O_2 band and the strong (9, 5) OH band at 10,014 A.

interest. A study of possible photochemical reactions by Bates and Nicolet (1950a, b) and by Herzberg (unpublished*) has shown that the most likely reaction is $O_3 + H = OH^* + O_2$ (3·3 ev. energy). This reaction has a large cross section, and although it involves *two* minor constituents, this is compensated by the fact that it is a two-body reaction. The energy available in this reaction is sufficient to excite the $v' = 9$ level, but insufficient to excite the $v' = 10$ level, which is in complete accordance with the absence of transitions from $v' = 10$, while transitions from $v' = 9$ are the most intense.

A radiative recombination was also considered by Herzberg (unpublished), wherein the oxygen and hydrogen atoms approach along the $^2\Sigma^-$ potential curve, which can have only a slight minimum. A radiative transition could then occur, allowing the OH molecule to arrive in the upper vibrational levels of the $^2\Pi$ state. This transition would give rise to an intense continuum on the long wavelength side of 11,800 A. in addition to the Meinel bands. Although this mechanism can explain the selective population of the OH vibrational levels, it is not possible to account for the observed intensities unless the amount of free hydrogen available in the upper atmosphere is considerably increased above the amount normally assumed.

Kaplan–Meinel Band of the Atmospheric System of O_2.

Kaplan (1947) discovered that the afterglow spectrum in pure O_2 consisted exclusively of bands from the $v' = 0$ level of the forbidden $^1\Sigma$–$^3\Sigma$ system of O_2. Although this work gave strong evidence that these emission bands might be present in the airglow, no sky observations were made. Independently, Meinel (1949, 1950 b) discovered that the (0, 1) O_2 band was a prominent feature of the infra-red spectrum of the night glow. In the case of the airglow, the (0, 0) emission band must be very intense in the stratum in which the radiation originates

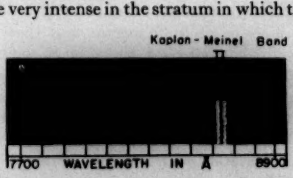


Figure 1. The infra-red night-glow spectrum showing the variability of the Kaplan–Meinel band.

since the (0, 0) band should be 27 times as intense as the (0, 1) band. The great mass of O_2 in the $v'' = 0$ level in the lower atmosphere, however, completely re-absorbs this radiation.

The fact that the laboratory and sky spectra are so similar indicates that a common mechanism affects the excitation. A primary excitation mechanism must have an energy of more than 1·72 ev. but definitely less than 1·90 ev. Bates (Private communication), however, has suggested that the selectivity could be produced by a secondary mechanism, wherein the electronic energy is exchanged during collisions according to the process:

$$O_2{}^1\Sigma(v' > 0) + O_2{}^3\Sigma(v'' = 0) = O_2{}^3\Sigma(v'' > 0) + O_2{}^1\Sigma(v' = 0)$$
$$O_2{}^1\Sigma(v' = 0) = O_2{}^3\Sigma + h\nu(0, n).$$

* Reported at the 1950 Pasadena symposium on Upper Atmosphere Research, sponsored by the Office of Naval Research.

This process would remove the restriction of a photochemical resonance phenomenon; however, the fact that this exchange process would work for the infra-red oxygen emission, but for none of the other electronic bands observed in the night glow spectrum makes the exchange suggestion questionable.

The Kaplan–Meinel band is a very variable feature. On some occasions the band has been enhanced as much as five times. The spectra in Figure 1 obtained during January 1950, show the OH bands in the region of λλ 7600–9000 A. with the Kaplan–Meinel band at the maximum and minimum observed intensities. Such variability will have to be accounted for by any excitation mechanism.

Telluric Absorption Bands.

Sommer (1932) identified numerous night-glow emissions in the red and near infra-red with the telluric absorption bands of O_2 and H_2O. With the feeble dispersion of the spectrograph employed by Sommer, it was impossible to distinguish by wavelength alone between the airglow emissions and the telluric bands. Spectra obtained by Meinel (1948) and Dufay (1949 a) with higher dispersion have shown that the telluric bands occur in absorption. The physical explanation of the absorption is quite clear; unless the emission originates very close to the observer, the preponderance of molecules in the ground state would re-absorb all emission from these bands. In the case of O_2, emission is observed from transitions to higher levels. There is no evidence, however, of emissions from H_2O to vibrational states above the (0, 0, 0) state. The near coincidence of some of the OH bands with the telluric absorption bands was responsible for the deception.

2.4. *Infra-Red Region beyond* 11,000 A.

The possibility of exploring the region of the night-glow spectrum beyond 11,000 A. is beset by many difficulties. On one hand, the presence of heavy telluric absorption bands due to H_2O and CO_2 will severely attenuate any emissions originating from the upper atmosphere. Regions of moderate transparency are few. In addition, detection methods for infra-red radiation are very insensitive in comparison with photographic materials and photoelectric cells in shorter wavelength regions.

Devaux (1931) reported observation of night-glow emissions in the region from 5 μ to 14 μ. These observations were made by means of a rock-salt prism and a very sensitive radiometer. The spectral resolution had to be very low, ranging between 0·5 μ and 1·0 μ bandwidth at 10 μ. Devaux reported feeble emissions near 5 μ and between 8 and 14 μ, with a secondary maximum at 10 μ, all of which were attributed to thermal emission from water vapour, ozone, and carbon dioxide. Adel (1949) has observed thermal emission from ozone in the day glow. However, no other observations have been reported of infra-red emission in the night glow.

The problem of accounting for emissions around 10 μ is not as difficult as in shorter wavelength regions. The amount of energy required for excitation can be supplied by the thermal velocities of the air molecules. In fact, heat emission due to the temperature of the gas should be quite prominent in spectral regions where the opacity is high. Any observable thermal emission, however, could not originate in the upper atmosphere because the opacity responsible for the emission will readily re-absorb the emission. The emission will therefore only be observable from the molecules in the immediate vicinity of the spectrograph. The only

possible upper-air emissions that could survive transmission to the ground would be fluorescent emissions to an excited state or emissions from a constituent not present in the lower atmosphere.

2.5. *Continuum*

Throughout the blue and violet regions the night-glow spectrum is dominated by a strong continuum upon which the Vegard–Kaplan bands are superimposed. It commences at a wavelength just below 4900 A. and terminates abruptly at 3900 A. Since this is the region in which the strong Vegard–Kaplan bands occur there is support for the suggestion by Barbier (1947 b) that the continuum is largely due to the superposition of the wings and weak bands of the Vegard–Kaplan system. Although this interpretation would explain qualitatively the abrupt increase of intensity at wavelengths above 3900 A., the absolute intensity of the Herzberg bands, for instance at 3546 A. and 3372 A., is considerably greater than any of the Vegard–Kaplan bands.

Other suggestions have been put forth concerning the origin of the blue continuum. Kastler (1946) suggested that it arises from the formation of negative oxygen atoms. Bates and Massey (1948), however, showed that this emission would be between 5000 A. and 6000 A., and could only account for a small fraction of the observed intensity. Bates and Massey (1948) suggested that the continuum is produced by radiative association between neutral atoms. While this mechanism could provide sufficient intensity, the resultant intensity distribution is unknown. The possibility that the continuum is produced by a polyatomic molecule should not be excluded. If it were emitted from heights of the order of 500 km. or greater (Barbier 1947) it would be difficult to find any mechanism.

The continuous background in the ultra-violet spectrum of the night glow is relatively weak. Barbier (1947 b) found that its intensity was approximately constant throughout the ultra-violet after correction for plate sensitivity and variation of dispersion.

An analysis of the light of the night sky in the blue and ultra-violet, according to Barbier (1947 b), gives the following results:

(1) 'resolved bands'	20%
(2) continuum	80%
(*a*) band wings (Barbier) .	53%
(*b*) zodiacal light (Dufay).	15%
(*c*) unresolved stars (Dufay) .	12%

The continuum is very weak in the green and yellow. An increase in intensity appears to be present in the infra-red; however, the great intensity of the Meinel bands makes the continuum difficult to observe. The observed continuum may, in fact, be due to the wings of these bands.

§3. SPECTRUM OF THE TWILIGHT GLOW

The spectrum of the twilight glow is so characteristically different from that of the night glow that it should be treated as a separate phenomenon. The twilight glow must represent the transition between a day glow and a night glow. The physical processes responsible for the night glow continue uninterrupted throughout the diurnal period. The day and twilight glows therefore constitute a superposition of strong emissions upon the night glow.

The solar radiation responsible for the strong emissions may be effective through several general mechanisms. The first mechanism would be *resonance scattering* by constituents that are rare in the lower atmosphere. For constituents that are abundant in the lower atmosphere a second mechanism of *fluorescent scattering* would alone be possible, since scattered resonance radiation from these constituents would be re-absorbed during transmission to the ground and the excitation energy lost by collisions. The two remaining possibilities would be the augmentation of the energy source in the form of *dissociation* or *ionization* energy. Of these four possible mechanisms, the first appears to be the primary mechanism for twilight-glow emissions in the visible region. There is some evidence, however, that fluorescent scattering may be present in the infra-red.

V. M. Slipher (1933) first noticed that the airglow spectrum at twilight showed the negative bands of N_2^+ for a short time. This observation was confirmed many years later by Elvey (1942) and by Vegard and Tonsberg (1942). The observations by Gauzit (1938), however, did not disclose the presence of these bands. The reality of the N_2^+ bands at twilight was proved by the work of Dufay and Dufay (1947) and by M. Dufay (1948). Dufay found that these bands are emitted between 60 km. and 150 km., with a maximum at 90 km. The distribution of intensity as given by Dufay is shown in Figure 2.

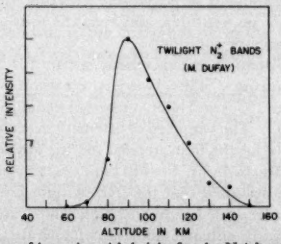


Figure 2. Distribution of intensity with height for the N_2^+ bands in the twilight glow. *After* Dufay.

Bates (1949 b) has made a study of the excitation mechanism for the N_2^+ bands at twilight and concluded that the bands were excited from the N_2^+ ground state by solar radiation. Swings (1949) has suggested an observational test of this mechanism. If the band emission is due to resonance scattering, then a high dispersion spectrum would show the effect of the solar absorption lines similar to the bands in cometary spectra. With a resolution of 1 A., the spectrum of the N_2^+ band at 3914 A. would show a sharp minima in the R branch near 3907·5 A., 3905·5 A. and 3901·9 A.

If the N_2^+ bands are emitted by resonance scattering, the presence of N_2^+ in the ground state must be explained. The transition probability, however, is so large for these bands that an extremely small equilibrium population will suffice. Bates (unpublished*) has estimated that as few as 100 N_2^+ atoms per cm^3 would suffice to produce the observed emission. Evidence concerning the origin of the N_2^+ in the ground state may be indicated by the observations by Dufay and Dufay (1947) that the intensity of the N_2^+ bands at twilight was

* Reported at the 1950 Pasadena symposium on Upper Atmosphere Research.

related to the magnetic activity and the appearance of aurorae in higher latitudes. This observation would appear to link the twilight N_2^+ bands to the N_2^+ bands of sunlit aurorae, the origin of the N_2^+ molecules being the same in both cases.

The most intense twilight glow feature is the Na D-line, discovered by Bernard (1938). The appearance of the Na D-line in emission can already be noted shortly after sunset. The observations by Bernard and numerous others seem to indicate that the Na D-lines are resonantly excited by solar radiation. Vegard and Kvifte (1945), on the other hand, cite observations that were interpreted as indicating that molecular dissociation was responsible for the emission. Penndorf (1950) also finds that molecular dissociation offers a satisfactory explanation of the Na D-lines during twilight.

The distribution of sodium atoms was found by Elvey and Farnsworth (1942) to be exponential between 70 km. and 115 km., following closely the total number of molecules. Apart from the sodium enhancement from below 130 km., there is a late twilight effect, discovered by Barbier and Roach (1950). Assuming that the late twilight enhancement is also due to resonance scattering, the emission originates in a region extending from 200 km. to 600 km. Barbier and Roach (1950) found that the variation of intensity with height indicated a scale height of 250 km., which corresponds to a kinetic temperature of 6000° K. They have suggested that this scale height, although it agrees closely with that for the 6300 [OI] enhancement (Elvey 1948), may be fictitious, and that the sodium concentration is due to accretion of interplanetary sodium.

A smaller, but well-defined, enhancement occurs for the red [OI] doublet in the twilight-glow spectrum. This twilight emission was first observed by Cabannes and Garrique (1936) and later studied in detail by Elvey and Farnsworth (1942) and Elvey (1948). The variation of intensity in early twilight is exponential and shows a larger scale height than normal for the atmosphere between 80 km. and 140 km. In the late twilight, $h > 140$ km., the rate of decrease in intensity diminishes. If it were assumed that the late twilight emissions were resonance scattering, then a very large scale height and high temperature would be necessary. The fact that the morning enhancement is considerably less than the evening enhancement points out that all of the enhancement cannot be due to resonance scattering. The constituents involved in the excitation process appear to be more abundant shortly after sunset than before dawn.

Bates (1948) has re-emphasized the importance of collisional de-excitation quenching on the forbidden emissions in the enhancement in the lower regions of the upper atmosphere. This suggestion, however, does not reach the core of the problem, since the larger scale heights are associated with great heights, where the collisional de-excitation for a lifetime of 96 seconds would be negligible. It is of prime importance, once again, that more observations be obtained under stricter conditions. The role of the continuum, in particular of the strong OH bands overlapping the 6300 A. [OI] emissions, should be examined. Twilight observations should provide an important tool for the investigation of the upper atmosphere once the interpretation is established.

Dufay and Dufay (1948) have reported a slight enhancement of the 5577 A. [OI] line in the twilight glow. The enhancement is about twice the night intensity. No confirmation has been reported as yet.

Elvey (1948) reported that the broad band at 6560 A. in the night-glow spectrum showed an enhancement at twilight. Later observational material by Elvey (1950), however, has failed to confirm the earlier observations.

On the whole the infra-red region of the night glow does not appear to be enhanced at twilight. Detection of this would be difficult because of the variability of the Meinel OH bands and the Kaplan–Meinel O_2 band. Meinel (unpublished) has recently studied the Kaplan–Meinel band and found evidence that this band is enhanced in the very early twilight.

§4. THE SPECTRUM OF THE AURORA

The auroral spectrum has been extensively observed, yet the analysis has been in an unsettled state for many years. The numerous observations made by Vegard and his colleagues are of great value; on the other hand, some of Vegard's analyses and often improbable identifications have led to scepticism and have caused important implications of some of the less intense auroral features to be ignored.

A new basis upon which to regroup the observational facts was provided by the analysis of the auroral spectrum by Bates, Massey and Pearse (1948) and by Swings (1949). In the last few years observations have confirmed the presence of OI and NI permitted lines and hydrogen lines, proposed originally by Vegard. Other faint features that had once been attributed to atomic emissions were identified as molecular bands. Swings and Nicolet focused attention on the fact that molecular bands from low temperature sources could easily be mistaken for atomic lines. In particular, the computation of the synthetic profile of the first Negative bands of O_2^+ by Nicolet (1948) enabled him to identify features previously held to be atomic.

The present discussion will in general be confined to auroral features of sufficient intensity to be recorded on moderate exposures. Spectra of the aurora in each spectral range are shown in Figure 3 (Plate *): spectrum (*a*), the blue and ultra-violet, was taken by Barbier and Williams at the University of Alaska with a one-prism quartz spectrograph; spectrum (*b*), the visible region, was taken by Gartlein at Cornell University, New York, with a two-prism flint spectrograph; spectrum (*c*), the infra-red region, was taken by Meinel at the Yerkes Observatory, Wisconsin, with a grating spectrograph.

4.1. *Ultra-Violet Region from* 3000 A. *to* 4000 A.

Second Positive Bands of N_2.

The most prominent features of the auroral spectrum to the violet of 3900 A. are due to the second Positive group of N_2, $C^3\Pi_u$–$B^3\Pi_g$. Bands from the sequences $\Delta v = -3, -2, -1, 0$, and $+1$ are present, of which the strongest are the (0, 1) and (0, 0) bands at 3577 A. and 3371 A., respectively. The auroral band intensities decrease more rapidly with increasing v' than those in laboratory spectra. The second Positive bands certainly present in the auroral spectrum are shown in Table 1.

Table 1. Second Positive Bands of N_2 in the Aurora

λ (A.)	*I*	v', v''	λ (A.)	*I*	v', v''
4059	5	0, 3	3577	8	0, 1
3998	6	1, 4	3537	7	1, 2
3943	4	2, 5	3371	8	0, 0
3805	7	0, 2	3339	2	1, 1
3755	6	1, 3	3309	1	2, 2
3710	4	2, 4	3161	7	1, 0
3671	1	3, 5	3140	4	2, 1

Negative Bands of N_2^+.

Apart from the strong (0, 0) band at 3914 A., the N_2^+ bands in the ultra-violet are rather weak. Some weak auroral features can be explained by the sequences $\Delta v = -1$ and -2.

Vegard–Kaplan Bands of N_2.

The Vegard–Kaplan bands have been assumed to be present in both the night glow and the aurora. The auroral spectrum shows bands that are weak or absent in the night-glow spectrum. The relative intensities of the bands common to both spectra, shown in Table 3, show that the intensities of bands arising from a common level are different. This difference between the night glow and auroral intensity distribution cannot be explained by different mechanisms. It adds one more reason why the identification of these bands in the airglow should be reviewed. The Vegard–Kaplan bands present in the auroral spectrum are given in Table 2.

Table 2. Vegard–Kaplan Bands of N_2 in the Aurora

λ (A.)	I	v', v''	λ (A.)	I	v', v''
4534	1	3, 15	3855	1	3, 13
4424	2	2, 14	3767	5	2, 12
4319	2	1, 13	3683	3	1, 11
4170	3	3, 14	3602	2	0, 10
4072	4	2, 13	3502	5	2, 11
3978	1	1, 12	3425	4	1, 10
4044	3	5, 15	3268	2	2, 10
3948	1	4, 14	3198	2	1, 9

Table 3. Intercomparison of Vegard–Kaplan Band Intensities in the Night Glow and Aurora

v' \ v''	9	10	11	12	13	14	15
0	—	2 : 0·5	—	—	—	—	—
1	2 : 0	4 : 1	3 : 0	1 : 0	2 : 3	—	—
2	—	2 : 0	5 : 1	5 : 2	4 : 5	2 : 5	—
3	—	—	—	—	1 : 0	3 : 5	1 : 0
4	—	—	—	—	—	1 : 3	0 : 4
5	—	—	—	—	—	—	3 : 0

Left-hand figure in each case is that for aurora, right-hand figure is that for night glow.

Other Band Systems.

Spectra obtained by Barbier and Williams (1950) showed some features that were identified as night-glow bands superimposed on the auroral spectrum. Several members of the Herzberg and the Barbier bands were identified. They showed no auroral enhancement.

4.2. *Visible Region from* 4000 A. *to* 7000 A.

Negative Bands of N_2^+.

Bands of the $^2\Sigma_u - {}^2\Sigma_g$ system of N_2^+ are prominent in the visible region of the auroral spectrum. In appearance, these bands show a single sharp head formed by the P branch, with a violet degradation formed by the R branch. The bands of this system are sufficiently intense to permit their study with high dispersion. Recently Vegard (1950) has resolved the rotational structure of the R branch, making possible a good determination of the effective rotational temperature.

The strongest bands of this system belong to the sequences $\Delta v = 0, -1$, and -2. Some weak features could be ascribed to the sequences $\Delta v = +1, -3$, and -4. Bands of the sequence $\Delta v = -1$ are difficult to identify because they are obscured by strong overlapping bands of the second Positive system of N_2. Nicolet (1948) has shown that the bands of the sequences $\Delta v = -3$ and $\Delta v = -4$ can explain a number of weaker features ascribed by Vegard to ionized atoms and molecular hydrogen. The identification of these sequences, however, is subject to confirmation; higher dispersion studies should be made to obtain a definite answer. If the features are molecular, then the higher dispersion will 'dilute' them as long as the band is not resolved; if they are atomic, the higher dispersion will not appreciably affect the intensities for a given exposure time. The N_2^+ bands that are certainly present in the auroral spectrum are shown in Table 4.

Table 4. Negative Bands of N_2^+ in the Aurora

λ (A.)	I	v', v''	λ(A.)	I	v', v''
4709	8	0, 2	3914	18	0, 0
4652	5	1, 3	3884	7	1, 1
4600	2	2, 4	3558	3	2, 2
4554	1	3, 5	3835	1	3, 3
4278	12	0, 1	3582	?	1, 0
4236	8	1, 2	3564	?	2, 1
4199	4	2, 3	3558	?	3, 2
4164	1	3, 4			

First Negative Bands of O_2^+.

Nicolet (1948) and Nicolet and Dogniaux (1950) demonstrated that some bands of the $^4\Sigma - {}^4\Pi$ first Negative system of O_2^+ are present in the auroral spectrum. This identification has recently been confirmed by high dispersion studies by Vegard (1950). All bands of this system, however, are rather weak; the strongest of these identified are the (2, 0) and (3, 1) bands covering the region from 5231 A. to 5295 A. The (0, 0) band may contribute to the auroral feature at 5977–6010 A. It is difficult, however, to identify other bands with this system since they occur in the region of the stronger first Positive bands of N_2.

First Positive Bands of N_2.

The $\Delta v = 3$ and $\Delta v = 4$ sequences of the $^3\Pi - {}^3\Sigma$ first Positive system of N_2 constitute rather strong features in the red region of the auroral spectrum. The intensity distribution appears to favour strongly the lower vibrational levels of the $^3\Pi$ state as compared with laboratory intensities.

4.3. *Infra-Red Region from* 7000 A. *to* 11,000 A.

First Positive Bands of N_2.

The most intense features of the entire auroral spectrum between 3000 A. and 9000 A. are the infra-red first Positive bands of N_2. The spectrum by Meinel, shown in Figure 3 (*c*), shows the $\Delta v = 1$, 2, and 3 sequences. The appearance of the (3, 1) band is modified by telluric O_2 absorption. The first Positive bands of N_2 observed in the auroral spectrum are shown in Table 5.

No photographic spectra of usable dispersion have yet been obtained of the infra-red auroral spectrum beyond 9000 A. It is certain, however, that the $\Delta v = 0$ sequence near 10,000 A. must be very intense.

New Band System of N_2^+.

Meinel (1950 i) has isolated a new band system in the infra-red auroral spectrum. Several of these bands are as intense as the first Positive bands of N_2. Their structure and wavelengths, however, clearly distinguish them from first Positive bands. The bands are characterized by two or more maxima. The violet edge of the first maximum is sharp, while the red edge of the last maxima shows a strong wing. Five or possibly six bands belong to this system as evidenced by their related structure. The analysis by Meinel (1950 i) has shown that this band system originates from the long-missing $A^2\Pi$ state of N_2^+. The bands observed in the aurora are shown in Table 6. The (2, 0) and (4, 2) bands are strongly blended with first Positive bands of N_2.

Table 5. First Positive Bands of N_2 in the Aurora

λ(A.)	*I*	v', v''	λ(A.)	*I*	v', v''
8912	int	1, 0	6788	8	4, 1
8722	int	2, 1	6704	7	5, 2
8542	6	3, 2	6624	4	6, 3
7753	30	2, 0	6543	2	7, 4
7626	20	3, 1	6469	1	8, 5
7503	12	4, 2	6395	1	9, 6
7387	8	5, 3	6185	3	4, 0
7273	5	6, 4	6127	3	5, 1
7164	2	7, 5	6070	4	6, 2
6875	6	3, 0	6014	2	7, 3
			5959	1	8, 4

Table 6. $A^2\Pi$–$X^2\Sigma$ Bands of N_2 in the Aurora

λ_{aur}	*I*	v', v''	λ_{aur}	*I*	v', v''
6858	4	2, 0	7828	40	1, 0
7050	7	3, 1	8057	20	2, 1
7250	4	4, 2	8298	6	3, 2

4.4. *Microwave Region*

Forsyth, Petrie and Currie (1949) have observed emission from auroral displays in the 3,000 Mc/s. region. Since the pulses were observed with the radar transmitter switched off, it seemed probable that the pulses were associated with the auroral disturbance. The pulses arrived in a random manner, occurring in bursts lasting for a small fraction of a second. Each individual pulse appeared to have separations of as little as 0·010 second. The investigation is continuing with the object of determining the pulse frequencies for different types and intensities of aurorae.

This discovery of auroral emission in the microwave region foreshadows a new phase in auroral research. The determination of the spectral distribution of the microwave noise would provide information of great value concerning the nature of auroral disturbances.

4.5. *Atomic Emission Features*

In view of the problems still attached to the identification of permitted atomic lines, the discussion of all wavelength regions has been combined. Recent observations with new equipment have done much to clarify the situation, largely through the exploration of the infra-red.

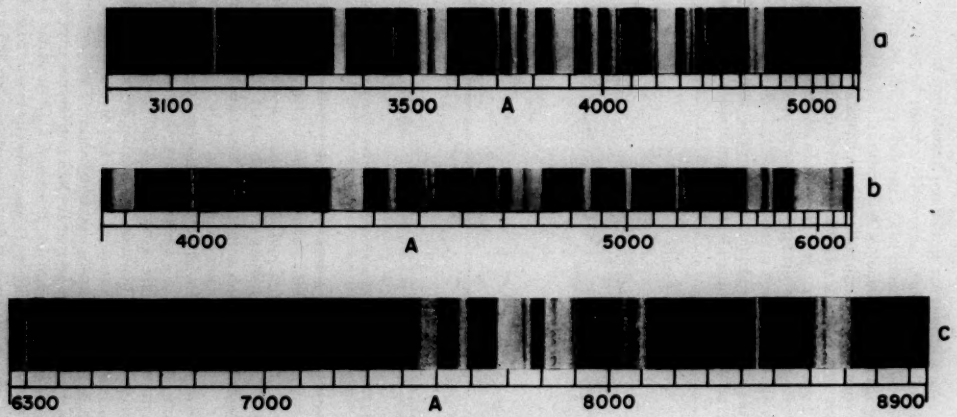


Figure 3. The auroral spectrum.

(*a*) Ultra-violet region: Barbier and Williams, $6^h\,15^m$, F/2 quartz 1-prism; (*b*) Visible region: Gartlein, $4^h\,20^m$, F/1·6 glass 2-prism; (*c*) Infra-red region: Meinel, $1^h\,30^m$, F/1 grating.

To face page 136

Hydrogen Balmer Lines.

The presence of members of the Balmer series of hydrogen is of fundamental importance in auroral theory. Hydrogen emission was first observed in some aurorae by Vegard and by Störmer. The presence of these emissions, however, was generally overlooked by others. A major step toward the general acceptance of the hydrogen identifications was through the extensive observations by Gartlein (1950). From observations covering a period of six years, he was able to show that the emissions ascribed to H_α, H_β, and H_γ showed the same relative intensities from aurora to aurora, throughout a wide range of intensities. The distinct broadness of the hydrogen lines, in comparison with other atomic lines provided the only uncertainty in this identification. Gartlein (1950), however, showed convincing evidence that the broadness could be attributed to the

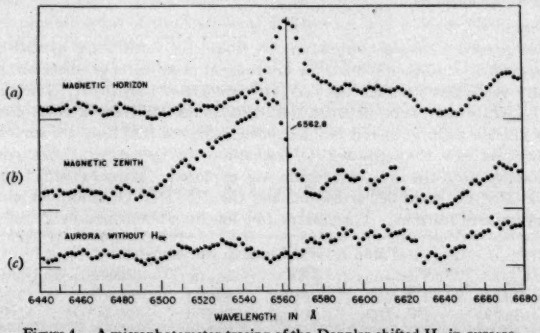


Figure 4. A microphotometer tracing of the Doppler-shifted H_α in aurorae. (*a*) Arch, magnetic horizon; (*b*) Arch, magnetic zenith; (*c*) Flames, magnetic horizon (H_α missing).

scattering of high velocity hydrogen atoms. Vegard (1950) has obtained observations that support this interpretation. Both Gartlein and Vegard, however, were unable to measure any significant Doppler shift for the broadened hydrogen lines. Vegard (1950) states that H_β is shifted, indicating a Doppler velocity, although his table of *measured* wavelengths does not indicate any shift greater than the probable error of the other measurements.

Recent observations by Meinel (1950 g) have led to an unambiguous evaluation of the nature of auroral hydrogen. Observations were made of the intense auroral storm of 18–20 August 1950 in the spectral region of H_α with a high resolution grating spectrograph. Spectra were obtained with the spectrograph pointed toward the *magnetic zenith* and toward the *north magnetic horizon*. Spectra of an auroral arch at the magnetic horizon show a strong, undisplaced, but broadened H_α. The line was symmetrical and undisturbed by the adjacent, but relatively weak, first Positive bands of N_2. Spectra of an auroral arch in the magnetic zenith, however, showed a very *asymmetrical* H_α profile. The violet wing of H_α was shifted 71 A., while the maximum was shifted approximately 10 A. A combined microphotometer trace is shown in Figure 4. The comparison auroral spectrum, Figure 4 (*c*), shows the first Positive bands without H_α. This spectrum was obtained from a flaming aurora the same night.

The asymmetrical displaced H_α profile shows conclusively that high velocity protons were entering the upper atmosphere, and that they were being sharply decelerated by it. The edge of the violet wing corresponds to a velocity of 3,200 km/sec. It is interesting to note that this velocity is considerably in excess of the velocities of 800–1,600 km/sec. deduced from the time delay for auroral storms. The higher velocity could be due either to an acceleration of the particles within a distance of several earth-radii of the earth or to a strongly curved trajectory of the particles between the sun and the earth. This observation leaves little doubt but that the auroral hydrogen is of solar origin, and that it represents one of the primary excitation sources for the aurora, as has been advocated by Swings (1948). It is of prime importance that extensive observations be undertaken to determine the distribution of hydrogen in different types and intensities of aurorae.

OI Lines.

The presence of the forbidden OI lines has been long established by interferometric studies. While they are present in all aurorae, there are marked intensity variations between the red [OI] and green [OI] lines. The intensity ratio I_{6300}/I_{5577} may vary between 0·05 and 10 for different auroral structure. In low latitude aurorae or red sunlit streamers the red [OI] lines are very intense.

The presence of permitted OI radiations in the auroral spectrum was questionable until the infra-red region was explored. Meinel (1948) found that the 3^3P–3^3S^0 OI multiplet at 8446 A. and the 3^5P–3^5S^0 OI multiplet at 7774 A. were prominent features. This observation has been confirmed by Petrie (1950).

Table 7. OI Lines in the Aurora

λ_{lab} (A.)	Transition	χ (ev.)	λ_{lab} (A.)	Transition	χ (ev.)
[6300·2]	2^3P_2–2^1D_2	1·96	4368·3	3^3S^0–4^3P	12·31
[6363·9]	2^3P_1–2^1D_2	1·96	7987·0	3^3P–$3s'^3D^0$	12·49
[5577·3]	2^1D_2–2^1S_0	4·17	6454·7	3^5P–5^5S^0	12·61
7774·7	3^5S^0–3^5P	10·69	7254·5	3^3P–5^3S^0	12·64
8446·4	3^3S^0–3^3P	10·94	6157·0	3^3P–4^3D^0	12·70
3947·5	3^3S^0–4^3P	12·23	3693·4	3^3S^0–5^3P	12·83

Barbier and Williams (1950) and Dufay (1949b) have taken this evidence and reconsidered the identification of permitted OI lines in the visible and ultra-violet. The identification of numerous OI emissions is now well established.

There appears to be a notable difference between the relative intensities in the auroral and laboratory intensities. The infra-red multiplets that arise from the lowest electron configuration above the ground configuration are considerably stronger. It appears that the excitation diminishes rapidly for excitation potentials above 12 ev.

The observed OI lines, listed according to energy levels, are given in Table 7. Other OI lines would be expected; however, their observation would be difficult, since they occur near strong molecular bands.

OII Lines.

A very narrow faint line, that is visible on long exposures, is found at 3727 A. by Barbier and Williams (1950) and Vegard (1946). It could be attributed to the 3727 A. [OII] ultra-violet doublet.

The spectrum of OII contains many lines in the visible and ultra-violet, so that numerous chance coincidences could occur. Vegard (1946) and Vegard and

Kvifte (1945) identify upwards of 50 lines, while Barbier and Williams (1950) find only four probable lines. The line at 4416 A., however, is the only feature of appreciable intensity in the aurora. The identification of any OII features, with the exception of the 3727 A. [OII] doublet, is therefore very doubtful.

OIII Lines.

Vegard has announced the identification of the [OIII] line at 5006·8 A. This identification was proposed for the strong auroral feature observed near 5004 A. by many observers. The line at 4959·0 A., which is a physical necessity if this identification is correct, has never been observed, although the intensity ratio is approximately the same as for the red [OI] lines. This identification must be rejected. No reliable identification of permitted OIII lines has been made.

NI Lines.

The identification of the [NI] lines at 3466·5 A. and 5197·8–5200·1 A. is well established. The 3466 A. [NI] line is clearly visible on all spectra covering this region with sufficient exposure. The 5200 A. [NI] line, because of its long lifetime, is observed only in low latitude aurorae or at great heights in streamers.

The identification of strong permitted NI lines in the infra-red auroral spectrum by Meinel (1948, 1950 b) has supported the identification of NI lines in other regions. As in the case of OI, the higher transitions are unusually weak in

Table 8. NI Lines in the Aurora

λ_{lab} (A.)	Transition	χ (ev.)	λ_{lab} (A.)	Transition	χ (ev.)
[5200·1]	$2p^4S^o$–$2p^3D^o$	2·37	8712·0	$3s^4P$–$3p^4D^o$	11·70
[5197·8]	$2p^4S^o$–$2p^3D^o$	2·37	8682·0	$3s^4P$–$3p^4D^o$	11·71
[3466·4]	$2p^4S^o$–$2p^3P^o$	3·56	8214·0	$3s^4P$–$3p^4P^o$	11·79
			4151·0	$3s^4P$–$4p^4S^o$	13·26

comparison with the infra-red lines. The relative intensities of the infra-red OI and NI lines have been used by Meinel (1950 b) and Petrie (1950) to estimate the ratio of NI/OI $\simeq 1$. This use of the infra-red NI lines is liable to error, since the lines are superimposed on the (2, 1) first Positive band of N_2.

Barbier and Williams (1950) could not identify any NI lines in the blue and ultra-violet, with the exception of 4151 A. and 4110 A. Some stronger lines of NI that would be expected are obscured by the molecular bands. These two lines, however, could also be attributed to N_2^+. Dufay (1949 b) lists eight possible NI lines of which four are unclassified.

NII Lines.

Ionized nitrogen presents a puzzling anomaly. Whereas NI lines are weak or absent in the visible region, several NII lines appear in surprising strength. Gartlein (1950) has suggested that the two auroral emissions at 5003 A. and 5676 A. are due to the $3p^3D - 3d^3F^o$ and $3s^3P^o - 3p^3D$ transitions in NII. The fact that these emissions maintain a constant intensity ratio of about unity strongly supports this identification. The widths of these auroral features is less than would be expected from the wide spread of wavelengths for the complete multiplets, indicating that the unusual intensity of these two emissions may be due to a fluorescent resonance cascade.

All other proposed identifications of NII emissions are weak auroral features, and are very doubtful.

Helium.

Vegard (1946) and Bernard (1948) have suggested the presence of He I in the auroral spectrum. Although a surprising number of coincidences can be found, the line at 3188 A., which should be very intense, is completely missing according to Barbier and Williams (1950). This observation, made under optimum conditions, appears to give sufficient grounds for rejecting this identification. In addition, consideration of the scarcity of helium at auroral heights, combined with the high excitation potentials, and also the small fraction of helium possible in the solar corpuscular radiation, affirm this rejection.

§5. HEIGHTS OF THE NIGHT-GLOW EMISSIONS

The heights from which each night-glow emission originates are very poorly known. The widely discordant results of different observers show that this aspect requires refinements in both technique and theory. With but a few exceptions, the reported heights are incompatible with the temperature determinations and possible excitation mechanisms.

The foremost method for determining the height of an emission stratum is due to van Rhijn. In the absence of any absorption or scattering in the lower atmosphere, the intensity of an optically thin homogeneous layer as a function of zenith distance is given by van Rhijn's equation. The difficulties in application arise from the patchiness of the emission and from the presence of a scattering and absorbing atmosphere below the emission stratum. In general, the lower atmosphere is stratified, with each layer having different optical properties. Furthermore, the background illumination from the zodiacal light and unresolved stars adds a contribution from $h = \infty$.

The evaluation of the roles of absorption and scattering is difficult. A small error in the evaluation of τ will cause large errors in the computed heights. Stratification, both in the ozone region and in the lower layers, complicates matters as does the fact that the scattering particle sizes and the scattering phase function are unknown. Barbier (1949) made a major advance by considering the effect of the scattering atmosphere, with the scattering represented by a spherical phase function. He found that a simple approximate formula

$$I(z) = I_0 e^{-\tau \sec z} + I_0 \Delta, \quad \text{(Barbier, 1st approx.)}\ \Delta = \tfrac{1}{2}(1 - e^{-\tau \sec z})$$

gave results that were in substantial agreement with a more rigorous treatment following the methods developed by Chandrasekhar (1944) for stellar atmospheres: (Chandrasekhar, 1st approx.)

$$\Delta = \frac{1}{\tau + 2/\sqrt{3}} \left\{ \cos z + \frac{1}{\sqrt{3}} - e^{-\tau \sec z} \left(\cos z + \frac{1}{\sqrt{3}} + \tau \right) \right\}.$$

Even with the more careful treatment of scattering and absorption, the experimental determination of τ is still difficult.

The most elaborate height determinations have been made by Roach and Barbier (1950) and by Roach and his colleagues. They have developed an automatic recording photoelectric photometer (Marlow and Pemberton 1949) that surveys the entire hemisphere of the sky, sweeping from horizon to horizon at numerous azimuths, covering the entire sky in a period of 30 minutes. During a single night as many as 20 complete sky surveys were obtained. The use of this instrument has largely eliminated the effects of the patchiness. The night-glow

emissions studied with this instrument were isolated by means of interference filters, having a pass-band of approximately 150 A. An attempt was made to correct for the continuum by having a separate interference filter located at 5200 A., which is a region in which no strong emissions are present.

The results thus obtained show a great improvement in the scatter of heights over results based on photographic or simple photoelectric methods. The scatter in height for the Na D-lines is shown in Figure 5. While the height of 110 km. for 5577 A. is in substantial accord with the proposed excitation mechanisms, the height of the Na D-line of 275 km. is too great for the emission to be due to photochemical processes (Bates and Nicolet 1950 c). If this is the actual height of the Na D-emission, the energy source and perhaps the sodium itself must be of extra-terrestrial origin.

The height determinations by Vassy and Vassy (1948), although elaborate, have led to enigmatic results. A photoelectric technique and ordinary filters were used, and the results were interpreted in terms of two layers, one at great

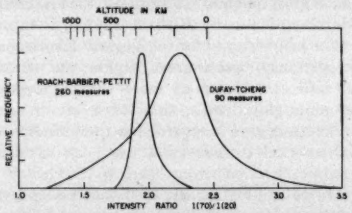


Figure 5. Intensity ratio scatter for photoelectric and photographic height determinations.

heights of the order of 1000 km. and another considerably weaker layer near 100 km. A range of height and relative intensity combinations could, however, explain their observations equally well. While it appears possible in this manner to isolate the effect of the continuum ($h = \infty$), actually the range of possible solutions is too great.

The photographic method of height determination used by Dufay and Tcheng (1948) shows a very large scatter in heights, although the average appears to be reasonable. The use of a spectrograph to isolate the radiations has the distinct advantage that the effect of the adjacent continuum is minimized. The height for the 5577 A. [OI] line of 80–100 km. agrees well with Roach and Barbier (1950), but the height for the Na D-line of 80 km. differs greatly. Theoretical considerations would make the height determination by Dufay and Tcheng preferable.

Roach, Pettit and Williams (1950) determined the height of the Meinel OH bands using simple photoelectric measurements. The reduction was simplified by the great intensity of these bands relative to any continuum and by the diminished absorption and scattering in the infra-red. Their determination gave a height of 70 km. for the OH emission. This value is in good agreement with the observed temperatures (Meinel 1950 e) and the excitation mechanism (Bates 1950). Meinel (1950 f) has used the height of the OH stratum to estimate the height of the Kaplan–Meinel band at 8645 A. Comparison of the zenith intensity distribution of the O_2 band with the OH bands on either side indicates that the O_2

emission also occurs near 70 km. Coupling this information with the temperature of the band, the most probable height of the O_2 emission is 80 km.

A determination of the height of the molecular bands in the blue and ultra-violet is difficult because of the strong continuum and the magnitude of the absorption and scattering in the lower atmosphere. Barbier (1947 b) has studied the Vegard–Kaplan and the Herzberg bands photographically. From a comparison of the intensities at zenith distances of 10° and 80°, he concluded that the Vegard–Kaplan emission originated at 900 km., while the Herzberg emission originated at 350 km. The range of heights varied between 5,000 km. and 200 km. A small uncertainty in the corrections could, however, drastically change these results.

The great height found for the Herzberg O_2 emission makes the reality of the determination doubtful. According to data given by Penndorf (1949) the ratio of oxygen in the molecular state to the atomic state at 350 km. would be of the order of 10^{-6}. No photochemical excitation process could account for the observed energy from N_2 at such great heights. In addition, the low temperature of 200° K. obtained for the Herzberg bands by Barbier (1947 b) is incompatible with the height of 350 km. The height found for the Vegard–Kaplan bands could be real, since molecular nitrogen may exist at great heights; the bands might be excited by extra-terrestrial sources. However, since the night-glow Vegard–Kaplan bands, as all other night-glow bands, show little or no enhancement during aurorae this height determination is regarded as questionable.

The trend of theoretical consideration has been to locate all night-glow emissions in the ' chemical kitchen ' from 70 km. to 120 km. While we should not try to force observations to fit a theory, it should be said that when striking contradictions occur between uncertain observations and plausible theory, one is forced toward the interpretation given by theory.

Perhaps the only way of obtaining reliable heights for the night-glow emissions under 200 km. is by rocket techniques. The emission could be detected even during the day, since ordinary scattering would be negligible above 70 km. The rocket should be fired at night, however, to avoid enhancement of emissions and resonance scattering of permitted transitions by solar radiation. Plans are now under way to carry out height determination by rocket observation for the 5577 A. [OI] and Na D-lines and for the Meinel OH bands.

Further theoretical development should be attempted in order to allow the accurate determination of the effective value of τ. A possible procedure would be to evaluate the equation of radiative transfer of stellar radiation through the atmosphere and also the equation of transfer of the night-glow emission in order to relate in a precise manner the role of τ in both equations. Observations of stars, either spectroscopically or with interference filters, could then lead to the evaluation of the effective τ. The necessary work has not yet been undertaken; however, it would not seem that any difficulty is presented, since the methods originated by Chandrasekhar (1950) could be applied directly.

§6. TEMPERATURE DETERMINATIONS FROM THE NIGHT-GLOW EMISSIONS

The temperature of an emitting source can, in principle, be determined from the spectral features. The dilute and often selective excitation of the night-glow emissions, however, immediately rules out temperature determinations based on populations in different atomic energy levels or molecular vibrational levels.

Only two remaining possibilities exist for the upper atmosphere: the first is by measuring the very small Doppler broadening of the atomic lines; the second is by measuring the rotational intensity distributions of the molecular bands.

While the determination of Doppler widths is difficult, the interpretation is unambiguous. The natural line width is readily determined by the transition probability of the emission. For the Na D-lines, the natural width is rather large, so that the Doppler component would be hard to detect. The forbidden 5577 A. [OI] line, on the other hand, will have a negligible natural width. In practice, the high resolving power required for such temperature determinations is difficult to obtain. The first effort along this line was by Babcock (1923), who found that the 5577 A. [OI] green line gave sharp interference fringes when the order of interference was 85,000. Angenheister (1932) and later Spitzer (1949) have pointed out that this order of interference is sufficient to give a useful temperature determination, since the corresponding upper limit of the temperature is about 900° K. If observations could be made with an order of interference of approximately 400,000, a useful temperature could be obtained.

The measurement of temperature by means of the rotational profiles of the molecular bands is generally employed. The validity of a rotational temperature as indicating the kinetic temperature is dependent on several factors. The most important factor is the lifetime of the excited state. This factor is important because chemical or other selective excitation mechanisms can lead to serious deviations between the kinetic and rotational temperatures. Such effects are well known in the laboratory. The fact, however, that the night-glow bands are due to forbidden transitions, and consequently have relatively long lifetimes, assures that the rotational distribution will be ' re-shuffled ' after excitation and prior to emission. The validity of the rotational temperatures obtained from night-glow emissions therefore depends on the atmospheric density and, correspondingly, on the altitude of the emitting region. If these emissions originate below the 300 km. level, the rotational temperature will be very close to the kinetic temperature. The situation can be quite different for permitted transitions, such as the N_2^+ bands in the aurora.

The temperature of the Vegard–Kaplan bands in the spectrum of the night glow obtained by Cabannes and Dufay (1946) of 230° K., and of the Herzberg bands obtained by Barbier (1947 b), of 200° K., were determined from low-dispersion profiles. Such determinations, however, necessarily have a low precision. In the case of the Kaplan–Meinel O_2 band, both the P and R branches are clearly observed. The resulting profile can be determined with more accuracy. Meinel (1950 f) has found a normal temperature of $150 \pm 10°$ K. for this band. Real variations of the temperature are found from night to night. During a strong enhancement of this band a temperature of 200° K. was observed. The lowest temperature that has been observed is 132° K.

The fact that the rotational structure of the Meinel OH bands can be observed directly has made it possible to determine very accurate temperatures for the OH stratum. Meinel (1950 e) obtained $260 \pm 5°$ K. from the rotational line intensities of the strong $^2\Pi_{3/2}$ component of the resolved P branches. A range of temperature of $\pm 15°$ K. is observed. The temperature of 260° K. and the height of 70 km. (Roach, Pettit and Williams 1950) are consistent with recent rocket data giving temperature with height.

The recent trend towards ascribing the airglow to the region below 120 km. makes temperature determinations of lesser importance, since the temperatures

are known in that region. If heights of this order of magnitude are confirmed then the temperature will at least lead to an accurate location of the emitting region, but the original hope that temperature determinations would probe the higher regions of the atmosphere will not be fulfilled. We must then turn to the study of the upper portion of auroral streamers and low latitude aurorae and radio data for a temperature probe above 150 km.

§7. TEMPERATURE DETERMINATIONS FROM THE AURORAL EMISSIONS

In the case of the aurora, the heights for a particular feature can be determined by relatively direct means. The transient nature of some of the interesting auroral structures, however, makes their study difficult, if not impossible. Nevertheless, auroral temperature determinations would be of value.

Auroral temperature determinations fall into two general classes: (i) low temperatures, as indicated by rotational structure intensities in molecular bands; (ii) high temperatures, as indicated by atomic line intensities and molecular vibrational intensities. The high temperatures are undoubtedly fictitious, since the particular intensities depend only on the excitation energy spectrum, which is certainly not characteristic of radiative thermodynamic equilibrium. The low rotational temperatures on the other hand are subject to the limitations previously described, since the auroral bands are largely permitted transitions. Again, however, a limitation is imposed on the use of permitted atomic and molecular emissions, since they arise from the lower regions in the aurora near 100 km.

In view of the foregoing considerations, the forbidden atomic lines present the best means of probing the temperature of the upper atmosphere. These lines extend to great heights in auroral rays and in the low latitude aurorae. Consequently, interferometric observations of sufficient resolution would be of great value.

A number of effective auroral temperature determinations have been made by Vegard and others. Vegard (1950) obtained temperatures of 219° K. for the N_2^+ bands, at an unknown height, from the resolved rotational line intensities. Petrie (1947, 1948, 1950), on the other hand reports temperatures from 3,000° K. to 10,000° K. from the relative intensities of permitted atomic lines. Rosseland and Steensholt (1933) also find temperatures between 2,000° K. and 7,700° K. from the relative molecular vibrational intensities. In view of these divergent results attention should be re-focused on the entire subject, with more emphasis on interferometric observations.

§8. CONCLUSIONS

Apart from the numerous problems pointed out under each separate topic, several problems of major importance still stand before us. The first is the state of the nitrogen in the upper atmosphere. This problem requires a re-examination of the Vegard–Kaplan bands in the night glow and of permitted NI lines in the aurora. Nicolet (private communication) has recently pointed out that any dissociation of nitrogen in or above the 300 km. region during an auroral disturbance will persist for a great length of time, so that sources of dissociation other than solar radiation should be considered in treating the equilibrium abundance of atomic nitrogen.

A second problem is the accurate determination of the emission heights of the airglow. Direct determinations during rocket flights would be of great

importance. However, further theoretical development along the lines suggested will probably permit accurate determinations from ground observations of the type first made by Roach and his colleagues.

A third problem is to probe the temperature of the atmosphere above 150 km. using the forbidden auroral radiations. It will be necessary to keep the exposure times short enough to catch auroral features for which a definite height can be obtained.

A fourth problem, of such scope that it was impossible to consider it in this discussion, is that of excitation mechanisms for the night-glow radiations and the aurora. The original hypothesis of Chapman (1931) that energy of dissociation provided an ample energy source for night glow emissions has again come into the limelight. The fact that numerous dissociation energies are poorly known, however, still poses an obstacle. Observational material is finally becoming available to allow the discrimination between primary and secondary excitation processes of the aurora.

In addition, many related problems remain. The ultimate objective will be the synthesis of the physical nature of the upper atmosphere based not only on the airglow and on auroral phenomena but on all other fields of upper atmosphere research.

ACKNOWLEDGMENTS

The author is exceedingly grateful to Dr. G. P. Kuiper and Mrs. A. B. Meinel for their co-operation in the preparation of this report.

REFERENCES

ADEL, A., 1949, *The Atmospheres of the Earth and Planets*, Ed. Kuiper (Chicago : University of Chicago), Chap. 10.

ANGENHEISTER, G., 1932, *Terr. Magn. Atmos. Elect.*, **37**, 431.

BABCOCK, H. D., 1923, *Astrophys. J.*, **57**, 209.

BABCOCK, H. W., 1939, *Publ. Astr. Soc. Pacif.*, **51**, 47.

BARBIER, D., 1945, *Ann. Géophys.*, 10, 224 ; 1947 a, *Ann. Astrophys.*, **10**, 47 ; 1947 b, *Ibid.*, **10**, 141 ; 1949, *Ann. Géophys.*, **5**, 83.

BARBIER, D., and ROACH, F. E., 1950, *Trans. Amer. Geophys. Un.*, 31, No. 1, 13.

BARBIER, D., and WILLIAMS, D. R., 1950, *J. Geophys. Res.*, in the press.

BATES, D. R., 1949 a, *Proc. Roy. Soc.* A, 196, 217; 1949 b, *Ibid.*, **196**, 562.

BATES, D. R., and MASSEY, H. S. W., 1948, *The Emission Spectra of the Night Sky and Aurorae*, Gassiot Committee Report (London : Physical Society), p. 48.

BATES, D. R., MASSEY, H. S. W., and PEARSE, R. W. B., 1948, *The Emission Spectra of the Night Sky and Aurorae*, Gassiot Committee Report (London : Physical Society), p. 97.

BATES, D. R., and NICOLET, M., 1950 a, *J. Geophys. Res.*, **55**, No. 3, 301; 1950 b, *C. R. Acad. Sci.*, Paris, **230**, 1943, 1950 c, *J. Geophys. Res.*, **55**, No. 3, 235.

BERNARD, R., 1938 a, *C. R. Acad. Sci.*, Paris, **206**, 448 ; 1938 b, *Ibid.*, **206**, 928 ; 1948, *The Emission Spectra of the Night Sky and Aurorae*, Gassiot Committee Report (London : Physical Society), p. 93.

CABANNES, J., and DUFAY, J., 1946, *Ann. Géophys.*, **2**, 290.

CABANNES, J., DUFAY, J., and DUFAY, M., 1950, *C. R. Acad. Sci., Paris*, **230**, 1233.

CABANNES, J., DUFAY, J., and GAUZIT, J., 1938, *C. R. Acad. Sci., Paris*, **206**, 870 and 1525.

CABANNES, J., and GARRIQUE, H., 1936, *C. R. Acad. Sci., Paris*, **203**, 484.

CAMPBELL, W. W., 1895, *Astrophys. J.*, **2**, 162.

CHANDRASEKHAR, S., 1944, *Astrophys. J.*, **100**, 76 ; 1950, *Radiative Transfer* (Oxford : University Press).

CHAPMAN, S. 1931, *Proc. Roy. Soc.* A., **132**, 353

DÉJARDIN, G., 1935, *Rev. Mod. Phys.*, **8**, 1; 1948, *The Emission Spectra of the Night Sky and Aurorae*, Gassiot Committee Report (London : Physical Society), p. 3.

DÉJARDIN, G., and BERNARD, R., 1938, *C. R. Acad. Sci., Paris*, **206**, 1947.

DEVAUX, J., 1931, *C. R. Acad. Sci., Paris*, **193,** 1207.

DUFAY, J., 1923, *C. R. Acad. Sci., Paris*, **176,** 1290 ; 1928, *Bull. Observ. Lyon*, **10,** No. 9; 1934, *C. R. Acad. Sci., Paris*, **198,** 107; 1941, *Ibid.*, **213,** 284 ; 1943, *Bull. Observ., Lyon*, **3,** No. 16; 1946, *Publ. Obs. Haute-Provence*, No. 3; 1949 a, *Ann. Géophys.*, **5,** No. 3, 183; 1949 b, *C. R. Acad. Sci., Paris*, **228,** 496; 1950, *Ibid.*, **231,** 000.

DUFAY, J., and DÉJARDIN, G., 1946, *Ann. Géophys.*, **2,** 249.

DUFAY, J., and DUFAY, M., 1947, *C. R. Acad. Sci., Paris*, **224,** 1834; 1948, *Ibid.*, **226,** 1208.

DUFAY, J., and TCHENG, M. L., 1942, *Cah. de Phys.*, **3,** No. 8, 1; 1948, *The Emission Spectra of the Night Sky and Aurorae*, Gassiot Committee Report (London : Physical Society), p. 62.

DUFAY, M., 1948, *C. R. Acad. Sci., Paris*, **227,** 777.

ELVEY, C. T., 1942, *Rev. Mod. Phys.*, **14,** 140 ; 1948, *The Emission Spectra of the Night Sky and Aurorae*, Gassiot Committee Report (London : Physical Society), p. 16; 1950, *Astrophys. J.*, **111,** 432.

ELVEY, C. T., and FARNSWORTH, A. H., 1942, *Astrophys. J.*, **96,** 451.

FORSYTH, P. A., PETRIE, W., and CURRIE, B. W., 1949, *Nature, Lond.*, **164,** 453.

FABRY, C., 1910, *C. R. Acad. Sci., Paris*, **150,** 272.

FABRY, C., DUFAY, J., and COJAN, J., 1934, *Étude de la Lumière du Fond du Ciel Nocturne* (Paris : Édition de la *Revue d'Optique*).

GARTLEIN, C. W., 1950, *Trans. Amer. Geophys. Un.*, **31,** No. 1, 7.

GAUZIT, J., 1938, *Ann. Astrophys.*, **1,** 334.

HERMANN, R., HERMANN, L., and GAUZIT, J., 1942, *Cah. de Phys.*, **12,** No. 1, 46.

HERZBERG, G., 1932, *Naturwissenschaften*, **20,** 577 ; 1949, *The Atmospheres of the Earth and Planets*, Ed. Kuiper (Chicago : University of Chicago), p. 193.

KAPLAN, J., 1935, *Publ. Astr. Soc. Pacif.*, **47,** 260 ; 1947, *Nature, Lond.*, **159,** 673.

KASTLER, A., 1946, *Ann. Géophys.*, **2,** 315.

KRASSOVSKY, V. J., 1949, *C. R. Acad. Sci., U.R.S.S.*, **66,** No. 1; 1950, **70,** *Ibid.*, No. 6.

KRON, G. E., 1950, *Publ. Astr. Soc. Pacific*, **62,** 264.

MARLOW, D., and PEMBERTON, J. C., 1949, *Rev. Sci. Instrum.*, **20,** 724.

MCLENNAN, J. C., and SHRUM, G. M., 1924, *Proc. Roy. Soc.* A, **106,** 138.

MEINEL, A. B., 1948, *Publ. Astr. Soc. Pacif.*, **60,** 357 ; 1949, *Symposium, Amer. Geophys. Un.* ; 1950 a, *Astrophys. J.*, **111,** 206 ; 1950 b, *Trans. Amer. Geophys. Un.*, **31,** No. 1, 9 ; 1950 c, *Astrophys. J.*, **111,** 432 ; 1950 d, *Ibid.*, **111,** 555 ; 1950 e, *Ibid.*, **112,** 120; 1950 f, *Ibid.*, **112,** 464; 1950 g, *Science*, **112,** 590; 1950 i, *Astrophys. J.*, **112,** 562; 1950 j, *C. R. Acad. Sci., Paris*, **231,** 1049.

NICOLET, M., 1948, *The Emission Spectra of the Night Sky and Aurorae*, Gassiot Committee Report (London : Physical Society), p. 105.

NICOLET, M., and DOGNIAUX, R., 1950, *J. Geophys. Res.*, **55,** No. 1, 21.

PENNDORF, R., 1949, *J. Geophys. Res.*, **54,** No. 1, 7 ; 1950, *Phys. Rev.*, **78,** 66.

PETRIE, W., 1947, *Canad. J. Res.* A, **25,** 293 ; 1948, *Ibid.*, A, **26,** 259 ; 1950, *J. Geophys. Res.*, **55,** No. 2, 143.

RAYLEIGH, Lord, 1920, *Nature, Lond.*, **106,** 8.

ROACH, F. E., and BARBIER, D., 1950, *Trans. Amer. Geophys. Un.*, **31,** No. 1, 7.

ROACH, F. E., PETTIT, H. B., and WILLIAMS, D. R., 1950, *J. Geophys. Res.*, **55,** No. 2, 183.

ROSSELAND, S., and STEENSHOLT, G., 1933, *Oslo Univ. Observ. Publ.*, No. 7.

SLIPHER, V. M., 1919, *Astrophys. J.*, **49,** 266 ; 1933, *Mon. Not. R. Astr. Soc.*, **93,** 666.

SOMMER, L. A., 1932, *Z. Phys.*, **77,** 374.

SPITZER, L., 1949, *The Atmospheres of the Earth and Planets*, Ed. Kuiper (Chicago : University of Chicago), p. 213.

STEBBINS, J., WHITFORD, A. E., and SWINGS, P., 1944, *Phys. Rev.*, **66,** 225 ; 1945, *Astrophys. J.*, **101,** 39.

SWINGS, P., 1943, *Astrophys. J.*, **97,** 72 ; 1948, *Publ. Astr. Soc. Pacif.*, **60,** 18 ; 1949, *The Atmospheres of the Earth and Planets*, Ed. Kuiper (Chicago : University of Chicago), p. 159.

VASSY, A., and VASSY, E., 1948, *The Emission Spectra of the Night Sky and Aurorae*, Gassiot Committee Report (London : Physical Society), p. 53.

VEGARD, L., 1946, *Physica*, **12,** 606 ; 1950, *C. R. Acad. Sci., Paris*, **230,** 1884.

VEGARD, L., and KVIFTE, G., 1945, *Geofys. Publ., Oslo*, **16,** No. 7.

VEGARD, L., and TONSBERG, E., 1935, *Z. Phys.*, **94,** 413 ; 1942, *Geofys. Publ., Oslo*, **13,** No. 1.

THE MICROPHYSICS OF CLOUDS

By B. J. MASON and F. H. LUDLAM
Department of Meteorology, Imperial College, London

CONTENTS

	PAGE
§ 1. Introduction	148
§ 2. Spontaneous nucleation in water vapour	149
§ 3. The initial stages of condensation—condensation nuclei	151
(i) Mode of action of a nucleus	151
(ii) Nature and origin of atmospheric condensation nuclei	153
§ 4. The growth of droplets in clouds and fog	159
(i) Factors influencing droplet growth	159
(ii) Determination of cloud droplet sizes	159
(*a*) Direct methods	159
(*b*) Indirect methods	160
(iii) Total liquid water content of clouds	161
(iv) Theoretical studies of the growth of droplet populations	161
(*a*) Schumann's 'growth by collision' theory	161
(*b*) Howell's computations	162
(*c*) Langmuir's 'time of rise' theory	164
§ 5. The initial stages of ice formation—ice-forming nuclei	166
(i) 'Freezing' and 'sublimation' nuclei	166
(ii) Laboratory experiments on the supercooling and freezing of water and aqueous solutions	167
(iii) Laboratory investigations into the properties and behaviour of ice-forming nuclei	169
(iv) Discussion of experimental results	175
§ 6. The growth of ice particles	177
(i) The formation of ice crystals and snowflakes	178
(ii) The formation and growth of hail	183
§ 7. The development of precipitation	184
(i) Forms of precipitation	184
(ii) The formation of precipitation elements	185
(iii) Growth of coagulation elements	185
(iv) The multiplication of raindrops by a chain reaction	187
(v) The artificial stimulation of precipitation	191
§ 8. Conclusion	192
References	194

ABSTRACT. The results of recent research into the microphysical processes involved in the formation of clouds and precipitation are described and assessed.

While the rate of spontaneous nucleation in water vapour is inappreciable until the supersaturation reaches about 400%, the presence of foreign nuclei in the atmosphere allows cloud formation with supersaturations of only about 0·1%. The mode of action, nature and possible origin of these atmospheric condensation nuclei are discussed. The growth of populations of these nuclei, to form cloud and fog, is considered in relation to the properties, concentration and size distribution of the nuclei and the properties of the air mass. The opposing theories of cloud-droplet growth by diffusion and by coagulation are critically examined. An account is given of the various techniques for determining drop-size distribution. The results of the many recent experiments on the supercooling and freezing of water and on the concentration and behaviour of ice-forming nuclei are examined. A new theory of the formation of ice crystals, snowflakes and hailstones is outlined. In dealing with the development of precipitation the conditions requisite for the production of drizzle, rain, hail and snow are discussed. The possible mechanisms of rain formation

receive special attention; it is suggested that in cumulus clouds, a water-drop chain reaction may be initiated by a small number of either wet hailstones or giant salt nuclei and that in tropical clouds the presence of the latter may render the intervention of ice crystals unnecessary. The possibilities of producing rain by artificial stimulation are briefly discussed in the light of our present knowledge of the microphysics of clouds.

§1. INTRODUCTION

RESEARCH in cloud physics aims at discovering the nature and relative importance of the physical processes occurring in natural clouds which culminate in the release of precipitation. Precipitation, which may reach the ground in the form of hail, rain or snow, is one of the most marked characteristics of 'weather' and represents the end product of a chain of events which form the main subject of this review. Although attention will be confined mainly to the *microphysical* processes, e.g. condensation, droplet growth and coalescence, ice crystal formation etc., it must be emphasized that the *macrophysics* of the problem is equally important. The factors which determine the large scale features of cloud development, e.g. the intensity, distribution and horizontal extent of vertical currents and the distribution of temperature and water vapour, decide the general trend of the microphysical processes and so must be taken fully into account if an integrated picture of the whole problem is to be obtained.

From the experimental standpoint we have two main methods of attack; observations may be made in the free atmosphere by using balloons and aircraft (with all their inherent difficulties), or we may design model experiments in the laboratory. It is extremely difficult however to simulate atmospheric conditions even approximately, so great caution must be exercised in interpreting the processes occurring in natural clouds on the basis of results obtained in the laboratory. A theoretical approach meets with equal difficulty; it is first necessary to ascertain by experiment and observation the dominant physical mechanisms and to assess the relative importance of the many complicating processes. Unfortunately there is still a paucity of reliable experimental and observational data. Nevertheless, much progress has been made during the last decade and we now have some knowledge of the main processes. Recent developments in this field pose many new problems not only for the meteorologist but for the experimental and theoretical physicist, the physical chemist and the crystallographer. Interest has been greatly stimulated by the possibility of inducing or inhibiting rain by the artificial seeding of clouds. Although success has so far been rather limited, the potentialities of this technique are becoming apparent and it may well be that when the conditions requisite for rain production are known more exactly, a greater degree of external control will be possible.

There is a continual interchange of the water substance between the surfaces of the earth and the oceans and the atmosphere. Water is continually evaporated, carried aloft, cooled by expansion of the air, condensed again into either the liquid or solid phase and eventually precipitated. It is this sequence of events which we shall discuss. Although the origin of thunderstorm electricity is very closely associated with the condensation processes in convective clouds, it is in itself a large and complex subject which cannot receive adequate attention within the scope of this article.

The purpose of this report is to take stock of the present position of the subject and to present a critical account of recent progress, both experimental

and theoretical. More attention will be given to those theories which appear to be reasonably well founded than to those of a more tentative nature, and an attempt will be made to emphasize the gaps in our knowledge which in the opinion of the authors merit priority in future research.

§2. SPONTANEOUS NUCLEATION IN WATER VAPOUR

Supercooled (supersaturated) water vapour is in a metastable state; in the course of time it therefore condenses into the liquid phase. In a homogeneous vapour the transition is effected by fluctuations whereby small aggregations of the liquid phase are continually produced; most of these aggregations disappear again, but if they are sufficiently large they are more stable than the original vapour phase and continue to grow, becoming *nuclei* for the growth of the liquid phase.

The critical size at which the chance aggregates are just in equilibrium with the metastable state, and above which they therefore serve as nuclei, may be calculated, and in the case of liquid aggregates in a supersaturated vapour is given by the well-known Thomson–Gibbs equation

$$r_c = \frac{2\sigma M}{\rho \boldsymbol{R} T \ln p/p_\infty}, \qquad \ldots\ldots(1)$$

where r_c is the critical radius of the liquid aggregate (assumed spherical), σ is the surface tension (specific surface energy) of the condensed phase, ρ is the density of the condensed phase, M is the molecular weight of the vapour, $\boldsymbol{R}$ is the gas-constant, T is the absolute temperature, p is the pressure of the supersaturated vapour and p_∞ is the saturated vapour pressure at temperature T over a plane surface of the condensed phase.

The methods of statistical thermodynamics also allow the calculation of the probability of the formation of aggregates of the critical size, and hence, in an extended phase, of the rate of formation of nuclei. Volmer showed this to be proportional to $\exp(-W_c/\boldsymbol{k}T)$, where $W_c = 4\pi r_c^2\sigma/3$ is the work of nucleus-formation and $\boldsymbol{k}$ is Boltzmann's constant. The complete expression for the rate of nucleus formation has been derived by Becker and Döring (1935) and may be written (Volmer 1939, p. 136)

$$J = \frac{ZXv}{r_c}\left(\frac{3W_c}{\pi \boldsymbol{k} T}\right)^{1/2} \exp(-W_c/\boldsymbol{k}T), \qquad \ldots\ldots(2)$$

where J is the rate of formation of nuclei/cm³ sec., Z is the number of individual molecules of the vapour per cm³, X is the number of the molecules striking 1 cm²/sec., multiplied by the condensation coefficient, and v is the volume occupied by a molecule in the liquid. In order to evaluate J this equation may be written

$$\log J = \log \frac{\alpha}{\boldsymbol{R}^2\rho}\left(\frac{2N^3M\sigma}{\pi}\right)^{1/2} + 2\log\frac{p_\infty}{T} + 2\log\frac{p}{p_\infty} - \frac{W_c}{2{\cdot}303\boldsymbol{k}T}, \qquad \ldots\ldots(3)$$

in which α is the condensation coefficient (assumed unity), and N is Avogadro's number. If $T = 260°$ K., then the values of J in vapour of various super-saturations are found to be as shown:

p/p_∞	1·01	1·1	2	3	4	5	6	7	8
$\log J$	−660,000	−7,600	−111	−29	−8·9	−0·2	4·75	8·0	10·1

From this it is clear that the rate of nucleus formation is inappreciable until the supersaturation reaches about 400%, increasing so rapidly at this stage that it is possible to speak of a critical supersaturation at which condensation first becomes observable. This result is in reasonable agreement with the evidence of experiments in which water vapour (freed as far as possible from impurities and ions) is subjected to sudden adiabatic expansion. No condensation is observed until, in the case of water vapour, expansion ratios of about 1·28 are used, corresponding to the brief production of saturation ratios of a little more than 5 at 260° K. The equation (2) also satisfactorily predicts the observed increase in the critical supersaturation with fall of temperature (Volmer and Flood 1934).

Becker and Döring (1935) have modified equation (2) to represent the formation of crystals of the solid phase directly from the vapour. In this modification, however, two factors of crucial importance occur whose numerical values are not known. The first of these is the σ-value for ice; the second is a geometrical factor related to the surface–volume ratio of the nucleus, and therefore dependent upon the crystal shape of the nuclear aggregate. Krastanow (1940) showed that the direct condensation of pure vapour into ice rather than into supercooled water cannot be expected until the temperature has fallen considerably below 0° C., and calculated that a critical temperature of about −70° C. represents the threshold at which this will occur; however, this calculated result is so sensitive to small changes in the values assumed for the two factors mentioned above that no confidence whatever can be attached to his value. Attempts have been made to discover a change in the relation between the critical supersaturation and temperature which might be associated with a transition from droplet to crystallet condensation (Sander and Damköhler 1943), but with inconclusive results, and indeed no reliable observations that water vapour ever condenses directly into ice without passing through the liquid phase have yet been made. Schaefer (1946) discovered that if a supercooled cloud of liquid droplets is momentarily chilled to a temperature below −39° C. enormous numbers of ice crystals are formed, and interpreted this as a critical temperature for the spontaneous condensation of vapour directly into ice. It seems more probable, however, that such a rapid chilling produces a momentary supersaturation high enough to cause spontaneous *droplet condensation*, and that the droplets quickly evaporate again unless their temperature lies below −39° C., when they freeze into crystals which even at the original temperature are in a supersaturated atmosphere, and which therefore survive and continue to grow. No attempt has yet been made to discover whether the critical temperature of −39° C. varies with the initial temperature of the supercooled cloud, or whether it remains constant, as must be expected if it represents the threshold at which spontaneous freezing of the liquid occurs.

So far it is the condensation in pure vapour which has been considered; if impurities are present in the vapour they may greatly facilitate the condensation. Thus if any solid motes are present at high humidities they may have films of adsorbed vapour upon their surfaces, and then behave as droplets of equal size, so that they represent ready-prepared aggregates which serve as nuclei for condensation at a supersaturation which diminishes with their size according to equation (1). Particles of hygroscopic substances hold liquid films even at relative humidities of less than 50%, and at higher humidities become droplets of solution, thus also representing efficient nuclei for condensation. In the atmosphere such impurities are present in such size and number that even the

most rapid cloud formation is associated with very small supersaturations, the value of the ratio p/p_∞ rarely exceeding 1·01, corresponding to a supersaturation of 1%. Nevertheless the behaviour of these atmospheric nuclei is of great interest and importance, and is discussed in some detail in the following paragraphs.

§3. THE INITIAL STAGES OF CONDENSATION—CONDENSATION NUCLEI

(i) *Mode of Action of a Nucleus*

The equilibrium vapour pressure p_r over the surface of a drop of pure water of radius r exceeds that over a plane water surface, p_∞, as expressed by equation (1): $\ln(p_r/p_\infty) = 2\sigma M/\rho RTr$. Thus a drop introduced into a saturated atmosphere, i.e. one in equilibrium with a plane water surface, will evaporate; a droplet cannot persist, much less grow, unless the environment is supersaturated. In the presence of solid particles whose surfaces are wettable droplet formation will be much facilitated; the larger the particles the smaller will be the supersaturation required. Continued condensation will occur on a solid insoluble nucleus of $r = 10^{-5}$ cm. if the supersaturation is only 1%.

If the drop is formed on a soluble nucleus the equilibrium vapour pressure at its surface is reduced by an amount proportional to the concentration of the solute. Salt nuclei are of great importance in atmospheric condensation. For this case Wright (1936) has derived the following equation:

$$\frac{p_r'}{p_\infty} = \exp(2\sigma M/\rho RTr) - \frac{750mC}{\pi\rho Wr^3}, \qquad \ldots\ldots(4)$$

where m is the mass of the solute of molecular weight W and C is a numerical factor depending on the nature and on the concentration of the solute. At equilibrium, with the surface temperature of the drop equal to that of the air, p_r' must be equal to the partial pressure e of the water vapour. But $H/100 = e/p_\infty$, where H is the relative humidity, so the condition for equilibrium becomes

$$\frac{H}{100} = \exp(P/r) - \frac{Q}{r^3}, \qquad \ldots\ldots(5)$$

where, $P = 2\sigma M/\rho RT$ and $Q = 750mC/\pi\rho W$. If P and Q are known, equation (5) can be used to determine the radius of droplets in equilibrium with air of a given relative humidity. Curves of H against r are plotted in Figure 1 for various values of Q, taking $T = 10°$ C. and $P = 11{\cdot}3 \times 10^{-8}$. At 100% relative humidity the depression of the equilibrium vapour pressure due to the presence of the dissolved salt is equal and opposite to the elevation caused by the curvature of the drop. If the humidity is further increased the droplet increases in size, remaining in equilibrium with the supersaturated atmosphere until a critical value of the supersaturation is reached corresponding to the maximum of the relevant curve in Figure 1. If thereafter, the supersaturation increases still further, the droplet will grow very rapidly since the larger it becomes the smaller is the relative humidity with which it is in equilibrium. The transition from an embryo droplet to a cloud particle of an order of magnitude larger is therefore quite sharp. If the supersaturation were maintained the drop would continue to grow very rapidly; in practice a stage is reached when the large numbers of rapidly growing droplets in the cloud extract water vapour faster than it is released by the cooling air and so the supersaturation is forced to retreat, and the growth rate decreases.

The uppermost curve of Figure 2 shows the supersaturations necessary for continued condensation on drops of pure water and applies equally to wettable non-hygroscopic nuclei. It is evident from the figure that a higher supersaturation

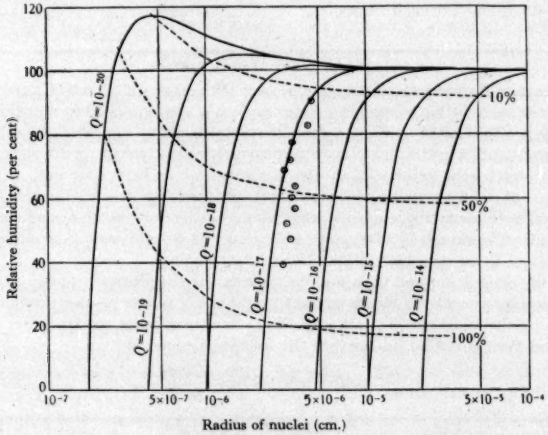


Figure 1. The equilibrium-radii of droplets of solution as a function of relative humidity (Wright.)

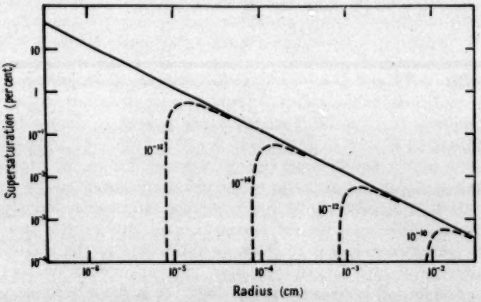


Figure 2. The equilibrium-radius of a droplet of pure water in a supersaturated vapour (solid line) and the corresponding equilibrium-radii of droplets containing the indicated masses (in gm.) of NaCl (broken lines).

is required to activate a non-hygroscopic nucleus than a hygroscopic nucleus having an equal radius at 100% relative humidity. The following table taken from the curves of Junge (1935) shows corresponding values of nuclear radius and the supersaturations required for condensation on droplets of H_2SO_4.

It is seen that all non-hygroscopic nuclei with r greater than $2{\cdot}3 \times 10^{-5}$ cm. and all hygroscopic nuclei with r greater than 8×10^{-6} cm. require supersaturations of less than 0·5% for them to act as centres of condensation.

Table 1. Radii of Nuclei which will promote Condensation at given Supersaturations

Supersaturation (%)	0·5	1	2	3	5	10	20
Non-hygroscopic (units 10^{-4} cm.)	23	11·5	6	4	2·5	1·2	0·65
Hygroscopic (r at 100% R.H.) (units 10^{-4} cm.)	8	4	2·5	1·7	1·1	0·65	0·4

Atmospheric nuclei are by no means of uniform size. As only small supersaturations are required for condensation to occur on large salt nuclei these will be activated first, and if present in sufficient concentration they may limit the rise in supersaturation and so prevent smaller or less efficient nuclei from becoming active. Thus the concentration of cloud droplets may be very much less than the total concentration of nuclei. The manner in which the concentration and size distribution of the nuclei influences the growth of the resultant cloud drops will be discussed in the next section.

(ii) *Nature and Origin of Atmospheric Condensation Nuclei*

While instruments of the Aitken 'dust counter' type may be used for counting the total number of nuclei in the air there is none which indicates the nature and size distribution of the nuclei—matters of much greater interest. The supersaturation produced in the Aitken counter is so great that nuclei down to ionic size become centres of condensation and so are counted. A large number of counts has been made with these instruments in many geographical positions, on land and sea, at sea level and in the upper air. Some measured concentrations of nuclei in various localities are shown in Table 2.

Table 2. Total Nucleus Concentration per cm^3 of Air in different Localities (*After* Simpson 1941a)

Locality	Places	No. of obs.	Average	Abs. max.	Abs. min.
City	28	2500	147,000	4,000,000	3500
Country (inland)	25	3500	9500	336,000	180
Ocean	21	600	940	39,800	2
Mountain (2000 m.)	25	190	950	27,000	6

There is an enormous range of particle concentration from $2/cm^3$ to $4 \times 10^6/cm^3$. The average concentration is smallest over the ocean and in the upper air and is increased by several orders of magnitude in air subject to industrial pollution.

The radii of atmospheric nuclei also cover a large range, from about 10^{-7} cm. for the small Wilson ions to about $10\,\mu$ for large sea salt nuclei, but there are usually sufficient nuclei with r greater than 10^{-5} cm. to prevent the activation of the smaller ones. It is these larger nuclei which are of greatest meteorological importance.

The mean size of atmospheric nuclei has been deduced by Wright (1939) from measurements of visibility. The visibility will depend on the nucleus concentration and the relative humidity; it will decrease if either of these factors increases, the other remaining constant. Wright analysed the observations of

visibility taken over a number of years both in clear and foggy conditions at various stations in the British Isles and calculated mean values of the atmospheric opacity over different ranges of relative humidity. From the available data he estimated the respective contributions to the opacity of the molecular scattering, absorption and scattering by solid particles, and scattering of light by water drops. Assuming that, over a long period, the average number of nuclei was independent of humidity he ascribed the changes of visibility to the change of nuclear size with humidity. At all stations there was a definite reduction of visibility with rise of humidity, confirming that some of the responsible nuclei were hygroscopic. Having found the relative change of mean nuclear size with relative humidity it was only necessary to find which of the family of curves shown in Figure 1 best fitted the observations to obtain the mean radius of the nuclei; the concentration was then determined from the relation between visibility and the size and concentration of the nuclei. In a later paper Wright (1940) applied the method in detail to visibility observations made over a period of five years at Valentia in S.W. Ireland. The curve of closest fit to the observations was found to be such that the mean radius of the nuclei at 78% relative humidity was about 5×10^{-5} cm., from which the concentration was calculated to be 63/cm^3. Wright reasonably assumed that the majority of these nuclei consisted of sea salt; the mass of salt contained in the average nucleus was then calculated to be $1{\cdot}1 \times 10^{-13}$ gm. These mean values may be compared with those obtained earlier by Köhler (1922), who analysed rime deposited from supercooled clouds for chlorides and estimated the mean drop size from measurements of coronae; he gives $r = 4 \times 10^{-5}$ cm. at 78% relative humidity and $m = 9{\cdot}18 \times 10^{-14}$ gm.

Wright found that the numerous nuclei of industrial areas were best represented by the curve of Figure 1 for which $Q = 10^{-17}$, with a corresponding radius of 4×10^{-6} cm. at 80% relative humidity. This analysis showed, however, that while they do not materially affect the visibility unless present in enormous numbers, the attendant smoke particles cause a marked reduction. He concluded, therefore, that the presence of sea salt nuclei is the dominant factor affecting visibility except in very smoky districts. It has often been suggested that combustion nuclei consist of sulphuric acid produced by oxidation of the sulphur impurities in the fuel. Sulphur dioxide is certainly a well-known constituent of town air and apparently combines with oxygen and water vapour to form sulphuric acid which is often present in appreciable concentrations.

A delightfully simple technique has recently been developed by Dessens (1946, 1947, 1949) for studying individual condensation nuclei. He has caught aerosol in country air on very fine spiders' threads (of about 10^{-6} cm. diameter) and examined it under a microscope; it consisted of small solid particles and small droplets of solution. Of the latter, a very few had radii of 1 μ to 10 μ but the majority were smaller than 1 μ at 60% relative humidity. The size distribution and concentration of these nuclei varied from day to day, the average concentration being of the order 100/cm^3. The droplets when placed in air of gradually decreasing relative humidity contracted slowly at first and then suddenly crystallized; the crystals from the larger drops could be observed under the microscope. Dessens observed that sometimes a single crystal was formed, sometimes a group of crystals stuck together and occasionally a number of separated crystals (Plate 1). This division may be a significant factor in increasing the number of salt nuclei in the atmosphere. Dessens reports that all the crystals he has examined are

cubical and concludes that they consist of NaCl; further support for this view was obtained from solubility and hygroscopicity tests. He finds no evidence for supposing that the nuclei consist of a mixture of sea salts. A nucleus of sodium chloride when subjected to a relative humidity of 78% becomes a droplet of saturated solution. At lower humidities, instead of crystallizing out as had been previously supposed, Dessens found that the particles tended to remain in equilibrium with the air as droplets of supersaturated solution and he was able to determine the radius of the drop as a function of relative humidity down to 40%, as shown in Figure 3. The same form of curve was obtained with

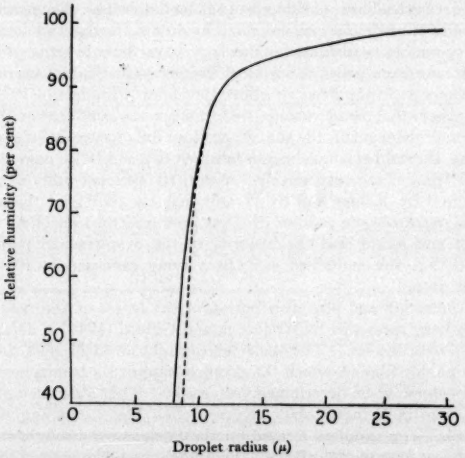


Figure 3. Experimental values of the radius of a natural (NaCl) or artificial ($ZnCl_2$) nucleus as a function of relative humidity (Dessens). ——— NaCl, - - - - $ZnCl_2$. *Reproduced from the Quarterly Journal by courtesy of the Royal Meteorological Society.*

artificial hygroscopic nuclei, e g. of $ZnCl_2$. A supersaturated drop remained stable for some time depending on the relative humidity, the breakdown of the supersaturation resembling a miniature explosion. Presumably such a drop will remain supersaturated until it becomes infected by a nucleus on which the salt crystals can form or until it becomes so supersaturated that it crystallizes spontaneously. As soon as crystallization begins the vapour pressure of the droplet becomes higher than that of the surrounding air; vaporization proceeds rapidly and may appear to be instantaneous. The separation of several crystals will be facilitated by this rapid evaporation.

Dessens' work is of considerable importance because it not only provides direct information on the nature of condensation nuclei but it throws new light on two serious objections raised by Simpson (1941 b) to Wright's thesis that sea salt is the most important source of condensation nuclei. Simpson argued that insufficient numbers of salt nuclei can be derived from sea spray to replace those

that would be lost in precipitation. From the annual global rainfall and the average size of cloud particles, he estimated that if sea salt nuclei were alone responsible for condensation, 50,000 salt nuclei would have to be produced every second from each square centimetre of the ocean surface subject to spray formation—which appears impossible. His estimate is made on the assumption that all precipitation is formed by the coagulation of cloud droplets—this, however, is not entirely justifiable for rain produced by melting snow, snowflakes being formed primarily by diffusion and not by collection of water droplets. This, together with the fact that the number of salt nuclei may be multiplied during crystallization of droplets as observed by Dessens, suggests that Simpson's figure may be a considerable overestimate. Crystallization of droplets will be induced by changes in humidity and the number of nuclei may be increased by orders of magnitude over a period of days. Favourable conditions for this may occur in convective clouds when nuclei which have been involved in cloud droplet formation are carried up and allowed to evaporate in the drier air above the cloud. - Findeisen (1937) has also attacked the theory that cloud particles are formed on sea salt nuclei on the grounds that the action of the wind on the sea can produce only relatively large and scarce drops of spray, the radii of which he puts between $10\,\mu$ and $100\,\mu$, containing about 10^{-10} and 10^{-6} gm. of salt respectively—an entirely different order of magnitude from that found by Köhler and by Wright, namely 10^{-13} gm. Köhler (1941) however, has measured the radii of droplets produced by a jet of air impinging on sea water and found that the majority of the droplets had radii between $0{\cdot}15\,\mu$ and $0{\cdot}35\,\mu$, the contained sea salt varying between 2×10^{-14} gm. and 10^{-12} gm. per drop.

The concentration and size distribution of the larger of the sea salt nuclei have recently been measured by Woodcock and Gifford (1949) at different levels up to 3,000 ft. over the sea. The nuclei were caught on small glass slides coated with a hydrophobic film on which the droplets assumed a hemispherical shape, allowing the volume to be determined very easily. They were then placed on a microscope stage in a thermostatically controlled box containing very dilute sulphuric acid. The droplets formed on the nuclei were brought into equilibrium with the air over the solution and the number and diameters of the droplets measured. Assuming these droplets to consist of concentrated sea water and having previously determined its vapour pressure, density and chlorinity, the average amount of chlorine per nucleus could be found. To test that the nuclei were actually sea salt they were compared with standard sea water nuclei produced from a spray; the curves of droplet radius against relative humidity were almost identical in the two cases. The chlorine content of the natural nuclei was also determined by titration, the results being in acceptable agreement with those obtained by the first method. The technique employed by Woodcock and Gifford did not allow them to catch nuclei smaller than those containing about 10^{-13} gm. of salt. Nuclei of this size were present in concentrations of about $10/\text{cm}^3$, but the concentration decreased rapidly for larger nuclei. An important feature of these observations however was the presence of a very few giant salt nuclei containing about 10^{-9} gm. of salt in extreme cases, which were present in concentrations of perhaps one per litre over the sea. Nuclei of this size would form droplets of radius $100\,\mu$ or more at 100% relative humidity (see Figure 2); such large nuclei could be important in the development of precipitation (see § 7). Woodcock and Gifford found that the total concentration of nuclei

decreased rapidly with increase in height when the atmosphere was thermally stable, but was reasonably constant up to 3,000 ft. under convective conditions.

Simpson based his second objection to the salt nucleus theory on the form of Wright's relative humidity–opacity curve which, he pointed out, was not consistent with the current belief that droplets of salt solution should crystallize when the relative humidity fell below about 78%. This latter assumption is disproved however by Dessens' observations, and if droplets can exist in the supersaturated condition down to relative humidities of 55%, their radii remaining almost constant, the shape of Wright's curve is immediately explicable.

Simpson, on the basis of some experiments by Junge, suggested that the condensation nuclei consisted of acid. Junge (1935) produced electrically charged nuclei by burning coal gas in a current of air and determined their size by measuring their mobility in an electric field. Curves of nuclear size as a function of relative humidity were obtained as shown in Figure 4: the radius decreases rapidly as the

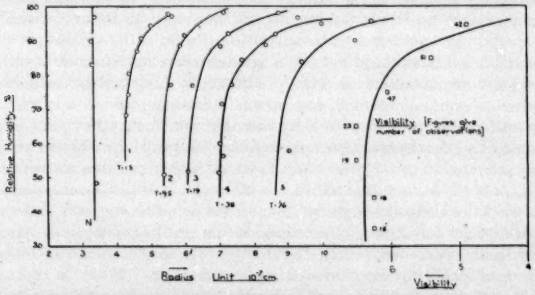


Figure 4. The radius of nuclei produced in a gas flame and the corresponding visibility as a function of relative humidity (Junge). *Reproduced from the Quarterly Journal by courtesy of the Royal Meteorological Society.*

humidity falls from 100% to 70% and thereafter remains constant. The corresponding curve of visibility against relative humidity has much the same form as that obtained by Wright (a fact to which Simpson attached considerable importance). Junge discussed the probable nature of these nuclei and concluded that they consisted of sulphuric acid formed from the sulphur in the coal gas.

The nature of combustion nuclei has also been investigated by Coste and Wright (1935) who compared the concentration of nuclei in an enclosed box before and after the burning of various fuels and the electrical heating of metallic surfaces. Large numbers of new nuclei were produced whatever the source of heat—even when pure absolute alcohol was burned and when a clean platinum surface was heated electrically in air previously freed from nuclei and SO_2. To explain their results Coste and Wright suggested that the nuclei so produced consisted of droplets of nitrous acid formed from the nitrogen, oxygen and water vapour of the air according to the following reactions:

$$N_2 + O_2 = 2NO \text{ (above } 620^\circ \text{ c.)},$$
$$2NO + O_2 = 2NO_2 \text{ (}620^\circ \text{ c. to } 140^\circ \text{ c.)},$$
$$NO + NO_2 + H_2O = 2HNO_2.$$

They were able to detect the presence of nitrous acid in their chamber after combustion had occurred, and from the simultaneous presence of the acid and the nuclei they suggested their identity. Further support comes from the fact that traces of nitrous acid have been detected in rainwater and the air of industrial regions. Coste and Wright also point out that nitrous acid may act as an oxidizing agent for sulphur dioxide and so produce sulphuric acid, and suggest that this is a more likely process than its formation from SO_3 in the gaseous phase.

So far we have considered only hygroscopic nuclei; there is also the possibility that non-hygroscopic particles may act as centres of condensation. Wigand (1913) measured the concentration of nuclei with an Aitken counter, then raised clouds of dust from coal, a beaten carpet, etc., and found that the number of nuclei detected was not appreciably increased. Similar results were obtained by Boylan (1926), and so for several years it was generally believed that condensation nuclei were all hygroscopic. In 1936 however, Junge showed that many kinds of non-hygroscopic dust can induce condensation, and intimated that the efficiency of such a particle did not depend markedly on the material of which it was made. It would indeed be surprising if solid non-hygroscopic particles with wettable surfaces could not act as condensation nuclei under a sufficiently high and prolonged supersaturation. Whether or not they will play a part in atmospheric condensation will depend on the concentration and size of the hygroscopic nuclei present. Non-hygroscopic nuclei may be quite important if there are only a few hygroscopic ones available, and will have a higher probability of being activated in convective clouds in which higher peak-supersaturations are reached.

The evidence then, although by no means conclusive, suggests that the most important nuclei for atmospheric condensation are hygroscopic in nature and are of two main types—salt nuclei derived from sea spray and maybe from desert regions, and acid nuclei produced by combustion. What is the relative importance of these two types? The larger and more efficient nuclei are almost certainly sea salt—in this respect the observations of Dessens and of Woodcock and Gifford appear conclusive. The question now arises whether there are sufficient salt nuclei to account for the observed concentrations of cloud droplets, which are of the order of a few hundred per cm^3. Dessens finds that the mean concentration of salt nuclei in country air remote from industrial areas is about $200/cm^3$; Woodcock's measured concentrations over the ocean are only of the order of $10/cm^3$, but he was able to capture only the larger ones, and as the numbers increase very rapidly with decreasing size the total concentration may well have been very much higher. Two thirds of the observations made over the ocean on the *Carnegie* (Shiratori 1934), gave total nucleus concentrations of less than $400/cm^3$, about one third of the counts showing between 100 and $200/cm^3$. More measurements of the *total* number of salt nuclei at different heights over the sea and over land in unpolluted air are required before we can answer this question satisfactorily. Comparison of the numbers occurring in the northern and southern hemispheres might prove informative because of the greater stretches of ocean and the stronger winds in the southern hemisphere. The chief difficulty is to design an apparatus which will catch nuclei down to 10^{-5} cm. diameter and allow the sampling volume to be determined. In the meantime, we cannot rule out the possibility of other hygroscopic nuclei playing a part. While nuclei of sulphuric acid may intensify fog in industrial areas it appears improbable that

nuclei produced by human activities are vitally important in producing cloud on a global scale. It is possible that nitrous acid is formed during lightning discharges and so helps to supplement the salt nuclei. The role of non-hygroscopic nuclei needs further investigation—it is by no means certain that they are entirely unimportant in droplet formation.

§4. THE GROWTH OF DROPLETS IN CLOUDS AND FOG

(i) *Factors Influencing Droplet Growth*

In the previous section attention was confined mainly to the nature and properties of the condensation nuclei and to the early stages of growth of an individual nucleus. We shall now consider the further growth of populations of nuclei, leading to the formation of cloud and fog. The growth rate of individual droplets depends not only on the hygroscopic and surface tension forces and the relative humidity as already discussed, but also on the processes of diffusion and thermal conduction which determine the rate of transfer of water vapour to and heat from the drop. Other very important factors are the concentration and size distribution of the nuclei, the rate of ascent of the air, its temperature, and the small scale turbulence in the cloud. The influence of these various factors on the growth of cloud particles has been partially investigated by theoretical methods and will be discussed later in this section, but first we shall describe some of the observations of drop size distribution and total liquid water content in clouds. These observations are of importance in that the drop size spectrum represents the result of several simultaneous physical processes and of the influence of the many factors just described, the relative importance of which can only be estimated by theoretical computations whose crucial test is the accurate prediction of the observed drop size spectrum. Both the total liquid water content and the drop size spectrum of a cloud are of fundamental importance in the development of precipitation and are very relevant to the problems of aircraft icing and visibility. A summary of recent experimental work has been given by Sheppard (1947).

(ii) *Determination of Cloud Droplet Sizes*

(*a*) *Direct methods.* Direct methods of determination of cloud droplet sizes involve the capture of cloud droplets on suitably prepared and exposed surfaces and the later counting under the microscope of the droplets or their impressions with reference to their diameters. This technique has been employed by Houghton and Radford (1938), Bricard (1939), Pauthenier and Brun (1940), Diem (1942) and May (1950). A glass slide coated with vaseline or magnesium oxide may be exposed to the cloud elements for a time sufficient to obtain an adequate sample, the deposit then being photographed and the diameters of the impressions measured. The accuracy of the method depends on obtaining a representative sample from the cloud or fog, preventing evaporation between sampling and counting, establishing the law of flattening of the drops in the film and avoiding shattering or coalescence of the captured drops. A detailed experimental and theoretical investigation of these matters by Langmuir (1944) shows that it is extremely difficult to obtain accurate size distribution curves and mean drop radii by this method; it is however capable of giving quite accurate values for particles down to 5 μ radius provided the fraction of the collecting surface covered by the drop is almost exactly 0·1. These techniques are of course

extremely laborious and considerable effort would be required to obtain representative samples in all types of clouds in different stages of their life history and under a variety of meteorological conditions. A recent development having some advantages involves the use in an aircraft of a special camera to photograph the droplets in a small sample of the free air.

The measurements published so far indicate that although the droplet radii in a given sample cover a considerable range, there is in general a fairly well marked single maximum in the drop radius spectrum and in the volume distribution curve, that is, the curve of nr^3 against r where n is the number of drops of radius r. The position of the maximum varies from one part of the cloud to another and is often quite different in different types of cloud. Houghton and Radford from 40 volume distribution curves obtained in sea fog found maxima at values ranging from 16 μ to 45 μ with a marked predominance between 20 μ and 25 μ, and about 30% of the total liquid water was contained in a 5 μ band centred on the individual maxima. It appears from the observations made so far that the mean drop radius of most clouds is less than for sea fog and is less in convective clouds than in layer clouds. The concentration of droplets is of the order a few hundred per cm^3.

(*b*) *Indirect methods.* The most important indirect method of droplet size determination consists in observing the nature and the amount of rime deposited on cylinders of varying diameter exposed to windborne supercooled cloud droplets. This method has been used on aircraft in Germany (Diem 1942) and in the U.S.A., and on Mt. Washington (Dolozel, Cunningham and Katz 1946). The results have been analysed by Langmuir (1944) and yield a kind of average drop radius.

When droplets in a stream of air approach a circular cylinder placed with its axis normal to the wind, they are deflected from their original paths by an amount depending mainly on the radii of the drops and the cylinder, and the wind speed. The catch of the cylinder increases with increasing drop radius and wind speed and with decrease in cylinder radius, according to the theory of Glauert (1940). Glauert's treatment has been modified by Langmuir to correct for the departure of the droplets from Stokes' law at high wind speeds. If the cylinder is fixed the deposition of rime is asymmetrical with respect to the cylinder axis and the theory soon loses its validity; this is overcome by rotating the cylinder to obtain a symmetrical catch. If now a number of rotating cylinders of different diameters are mounted side by side, the differing catch per unit width of the respective cylinders enables the total water content and the volume-median diameter of the droplets to be deduced.

Langmuir has analysed riming data obtained on Mt. Washington (elevation 6,290 ft.) in wind speeds ranging from 24 to 100 m.p.h., in temperatures from $-25°$ F. to $+29°$ F. and in clouds whose liquid water contents varied from 0·09 gm/m^3 to 1·13 gm/m^3, volume-median drop diameters from 6·4 μ to 24 μ and drop concentrations from 47/cm^3 to 1,560/cm^3. He found the volume-median drop diameter d_{vm} (in μ) to be related to the air temperature T (° F.), water content W (gm/m^3) and wind speed V (m.p.h.) by

$$d_{vm} = 0{\cdot}015(74 + T)\left(\frac{1000W}{V}\right)^{0{\cdot}40} \qquad \ldots\ldots(6)$$

the experimental data agreeing well with this formula in respect of W/V, a source of large variation in d_{vm}, but rather poorly in respect of T, a source of much

smaller variation in d_{vm}. The increase of d_{vm} with T and W is regarded by Langmuir as an important physical result which he goes on to relate with the process of droplet growth, as we shall see. The wind speed was regarded as a measure of the vertical velocity of the air in the cloud moving up the slope of the mountain and therefore as important in determining the age of cloud droplets. Langmuir considers the drop size to be very largely determined by this age. The number n of drops per cm^3 deduced from d_{vm} and W is given by Langmuir as

$$n = 0{\cdot}22(55 - T)V(10W)^{-0{\cdot}23} \quad \ldots\ldots(7)$$

or

$$n = 0{\cdot}32(70 - T)V\,d_{vm}^{-0{\cdot}70}. \quad \ldots\ldots(8)$$

The equations fit the data for n and V quite well, but the correlation between n and W (or n and d) and between n and T is poorer.

(iii) *Total Liquid Water Content of Clouds*

The total liquid water content of a cloud, however dispersed, is of considerable meteorological importance. A knowledge of its magnitude and distribution in various types of cloud would enable more precise calculations on the rate of growth of precipitation elements to be made. It enters also into the theory of cloud development, particularly of convective (cumulus) clouds, where processes causing changes of liquid water content release considerable quantities of energy.

It can of course be calculated from the drop-size distribution curves taken from measured volumes of cloud. The method however is inaccurate because of the difficulties attending the determination of drop-size distributions. The riming cylinder method will, as we have seen, yield the liquid water content but its use is restricted to supercooled clouds. A direct method of universal application would be of considerable value but so far no really satisfactory instrument has been produced, and so we still lack representative values of water content over a range of meteorological conditions in different types of cloud. One promising instrument is the capillary collector designed by Vonnegut (1949 a) for use on aircraft. The collector head has a porous surface of sintered metal with pores of 30 μ diameter; droplets colliding with the surface are drawn into a measuring system, their volume being measured by the displacement of water along a capillary tube. This instrument, although probably satisfactory in precipitating clouds, appears to miss a large fraction of the smaller cloud droplets, the efficiency of catch at 200 m.p.h. being only 0·75 for droplets less than 5 μ radius and only 0·5 for those of 3 μ radius.

(iv) *Theoretical Studies of the Growth of Droplet Populations*

(*a*) *Schumann's 'growth by collision' theory.* Schumann (1940) assumes that the dominant physical process in the growth of cloud droplets is the collision and resultant coagulation of droplets by virtue of Brownian motion, small scale turbulence and differential rates of fall under gravity. In developing this theory he makes two simplifying assumptions—that every collision results in coalescence, and that the frequency of collision is independent of the size of the drops—both of which are perhaps unrealistic. Assuming an initial drop size spectrum, Schumann derives expressions for the form of the drop-radius distribution and volume distribution curves at any subsequent time. These equations are compared with Houghton and Radford's three published distribution curves, there being excellent agreement with one of them and somewhat poorer agreement

in the two other cases. Schumann then proceeds to investigate whether coalescence by collision adequately accounts for the actual size of the drops. Having shown first of all that collision due to Brownian motion is unimportant compared with collision due to differences in terminal velocity if the 'mean' drop radius is greater than 3μ, he arrives at the following equations for this as a function of time:

$$\left.\begin{aligned} r_m &= 0{\cdot}0025(vt)^{0{\cdot}4} && r_m < 3\mu \\ \text{and} \quad r_m &= \frac{r_0}{1-3{\cdot}23\times 10^5 vt\, r_0} && r_m > 3\mu \end{aligned}\right\} \qquad \ldots\ldots(9)$$

where $r_m = r_0$ at $t = 0$, r_m being the radius at which the volume distribution curve reaches a maximum, v the total volume of water per cm³ of air, and t the time. Combining the two laws of growth and assuming that $v = 1$ cm³/m³, the times of growth from negligibly small droplets up to a maximum of 15μ radius (limit of Stokes' law) were calculated. These are plotted in Figure 5. The times required are excessive, especially for growth up to about 8μ, bearing in mind also that the water content of layer clouds is often less than the value assumed by Schumann. It appears then that coalescence by collision is not the dominant factor in the early stages of cloud droplet growth.

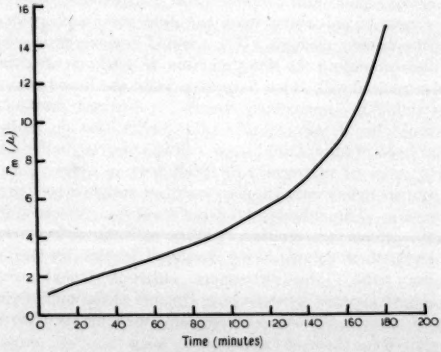


Figure 5. The mean radius of cloud droplets as a function of time as predicted by Schumann's 'growth by collision' theory.

(*b*) *Howell's computations.* Recently Howell (1949) has undertaken a detailed study of the growth of cloud drops in air cooled at a uniform rate, with the object of predicting the change of drop size spectrum with time. He assumes the droplets to grow by diffusion of water vapour to their surfaces, the coagulation of droplets being of negligible importance. He also takes into account the effect of the hygroscopic nucleus, the heat balance of the drop and the variation of vapour pressure over a drop with its temperature. He sets up the equation of growth in differential form and then proceeds by stepwise integration. By assuming that the ambient temperature and pressure remain sensibly constant during the cloud forming process, the number of independent variables is reduced to four: the relative humidity, the mass of the parent nucleus, the drop radius and the

time. As a starting point for a particular calculation, a concentration and a simple size distribution of nuclei are assumed and the subsequent march of the supersaturation and of the drop size distribution is computed for different rates of cooling. Use is made of the concept of total supersaturation, defined as the supersaturation that would exist if all the liquid water in the air were vapour at the same temperature. It is thus related to the concentration and size of the drops, imposing one of the conditions which must be satisfied at each step of the integration. Assuming that the total specific humidity of an air parcel remains constant during cloud formation, the rate of increase of the total supersaturation under assumed conditions of temperature and pressure is easily related to the rate of cooling of the air.

Howell's computations show that a group of droplets originating from nuclei of uniform size undergoes four distinct phases of growth: (i) before the supersaturation reaches a critical value corresponding to the maximum of the relevant curve of Figure 1, the drops grow slowly and do not lag far behind the equilibrium size except in the case of large nuclei with high rates of ascent; (ii) when the critical supersaturation is passed, rapid growth sets in and is terminated when the increased condensation on the rapidly expanding liquid surfaces prevents further rise in the supersaturation; (iii) the rate of growth approaches equilibrium with the rate of supply of vapour; (iv) the final phase, when the effects of the nucleus have become negligible.

Turning to more realistic conditions Howell investigates the subsequent growth of a reasonable concentration and size distribution of salt nuclei under conditions of rapid, moderate and very low, steady rates of ascent. His initial conditions and results are shown in Table 3.

Table 3. The Growth of non-uniform Groups of Nuclei under various atmospheric Conditions

Rate of ascent (cm. sec.$^{-1}$)	400	30	1·5
Total nucleus concentration (per cm^3)	500	2000	667
Range of nuclear masses (moles)	10^{-19} to 10^{-15}	10^{-16} to $10^{-13\cdot8}$	10^{-16} to $10^{-13\cdot5}$
Peak supersaturation (%)	0·36	0·072	0·056
Time to attain peak supersaturation (sec.)	25	70	400
Mass range of activated nuclei (moles)	10^{-17} to 10^{-15}	$10^{-15\cdot62}$ to $10^{-13\cdot8}$	$10^{-15\cdot5}$ to $10^{-13\cdot5}$
Concentration of activated nuclei (per cm^3)	250	410	40
Radius at peak supersaturation (μ)	1·5 to 3	1·5 to 4·5	2 to 8
Time taken to grow to $r=5\mu$ (sec.)	150–100	2200–100	2450–30
Modulus of size distribution (m)* at 100 m. above cloud base	0·06	0·33	0·23
Volume-median radius at 100 m. above cloud base	$\sim 3\mu$	$\sim 5{\cdot}5\mu$	$\sim 14\mu$

* $m=\sqrt{2}\sigma/r_{vm}$ where r_{vm} = volume-median radius and σ = standard deviation of drop volumes. A small value of m denotes a more nearly uniform spectrum.

We may draw the following conclusions from Howell's calculations. Higher rates of cooling produce larger peak supersaturations and cause a larger proportion of the smaller nuclei to be activated. Condensation sets in first on the larger nuclei; if they are present only in small numbers, and so are unable to assimilate the vapour as fast as it is released by the cooling air, the supersaturation rises and allows the smaller nuclei to be effective. Moreover, the great majority of the nuclei which are activated grow to the critical size and become droplets; the maximum concentration of drops coincides with the peak supersaturation. Thereafter the supersaturation falls because vapour diffuses to the many condensation centres more quickly than it can be replaced by the cooling air, and now

the smallest drops start to evaporate. These droplets are less than 0·1 μ radius, however, and so contribute very little to the growth of the larger drops. The most uniform drop spectrum was produced in the case of relatively few nuclei in rapidly expanding air. In all three cases the drop size distribution becomes more nearly uniform with time, the smaller droplets growing faster relative to the larger ones. The small modulus of distribution produced even by slowly rising air indicates that a quite broad spectrum of nucleus sizes cannot directly produce a marked degree of heterogeneity in a cloud formed in uniformly cooled air. By comparing his calculated drop size spectra with those actually measured on Mt. Washington and in flight, Howell concludes that the narrow spectra characteristic of young clouds may be adequately explained by his equation of growth applied to uniformly cooled air. In more inhomogeneous clouds, however, other factors must modify the drop size spectrum, e.g. turbulent mixing of dissimilar air parcels containing drops of different size and temperature and the fact that the cloud will be formed by parcels of air rising at different speeds rather than by a whole air mass moving bodily at uniform vertical velocity. Howell's calculations also show that in clouds containing a few hundred nuclei per cm^3 the supersaturation attains a peak value of only about 0·1% and can surpass 1% only under extreme circumstances.

(*c*) *Langmuir's 'time of rise' theory.* The rates of growth obtained by Howell are considerably less than those obtained earlier by Langmuir (1944), who considered that the most important factor in the growth of cloud droplets up to a radius of about 15 μ under near-saturation conditions was the difference in vapour pressure over drops of different sizes leading to evaporation of the smaller drops and condensation on the larger, with a concurrent decrease in drop concentration. Langmuir calculated the growth-rate of a cloud of droplets on the assumption that at any given instant the vapour pressure was in equilibrium with the droplets of *average size*, the smaller ones tending to evaporate and the larger ones to grow, the average size increasingly progressively with time. He also assumed that the *relative* size distribution function remained unchanged during this process, the fraction of drops disappearing in unit time therefore varying inversely as the life t of the drops of average size, t representing the time taken for the air to rise from the base of the cloud to any level under consideration.

Langmuir arrives at the following equation for the mean radius r of the drops after a time t:

$$r^3 = r_1^3 t/t_1 + (t/k\beta)\ln t/t_1, \qquad \ldots\ldots(10)$$

where $r = r_1 = 10^{-5}$ cm. at $t = t_1$, and $t_1 = 5 \times 10^{-11} k\alpha$. k is a constant of order unity depending on the size distribution function,

$$\alpha = \frac{1}{2}\left(\frac{\pi}{2}\right)^{1/2}\left(\frac{RT}{M}\right)^{3/2}\frac{p^2}{\sigma p}, \qquad \beta = \frac{1}{6D\sigma p}\left(\rho\frac{RT}{M}\right)^2,$$

where D is the coefficient of diffusion of water vapour in air and the other symbols are as defined in previous sections.

Table 4 shows calculated values of the droplet mean radius (in microns) as a function of time and temperature, assuming $k = 1$.

These figures may be compared with those of Schumann and of Howell. On the growth by collision theory a droplet takes 110 minutes to grow to a radius of 5 μ; from his diffusion theory Howell predicts times varying from about 100 to 2,000 sec., depending on the size of the nucleus and the rate of ascent, while Langmuir's values vary from 10 to 120 sec. depending on the temperature.

In order to apply his theory to the rime measurements on Mt. Washington, Langmuir had to determine t in each case. He took $t = h/sV$, where h is the height of the level of observation above the cloud base, s the slope of the mountain ($=0{\cdot}32$) and V the wind speed. The height h was determined from the water content W of the cloud and the temperature at the level of observation. Finally, Langmuir obtained the equation $t = f(T)W/V$. Having found t, the theoretical value of the radius was calculated from equation (10), the value of k being decided by setting k times the mean theoretical value of r^3 equal to the mean observed value of r^3; this gave $1/k^{1/3} = 1{\cdot}25 (k = 0{\cdot}51)$. Excellent agreement was claimed between the theoretical and observed values of r, while the empirical relations deduced from the riming data (equations (6), (7), (8)) were also in good agreement with the theory.

In spite of this apparent agreement it appears from what we have said already about the physical processes of droplet growth that the mechanism postulated by Langmuir cannot be of major importance in determining the growth rate of droplets in a rising air current in which vapour is continually being made available

Table 4. Radius of Water Droplets (in μ) as a Function of Time and Temperature

T(°c.)	t_0 (sec.)									t_1(sec.)
	1	10	40	80	120	300	600	1200	3600	
20	2·31	5·48	9·13	11·75	13·60	18·9	24·2	31·0	45·9	0·0009
10	1·83	4·35	7·29	9·38	10·86	15·16	19·4	24·9	36·8	0·0016
0	1·42	3·42	5·70	7·35	8·55	11·9	15·2	19·4	29·1	0·0030
−10	1·07	2·60	4·37	5·65	6·56	9·17	11·8	15·2	22·6	0·0060
−20	0·78	1·93	3·26	4·24	4·90	6·86	8·9	11·4	17·0	0·0127

Values are based on $k=1$; for other k multiply by $1/k^{1/3}$.

for condensation. Under these conditions no droplets will evaporate until the supersaturation, having reached its peak, commences to fall; only then can the Langmuir process begin to operate. Howell's calculations show moreover that the ultimate droplet concentration is very largely determined before this stage is reached; the number and size of the droplets which evaporate is very small and so they are not important in determining the subsequent form of the drop-size spectrum. If, however, the rate of cooling then decreases or if the relative humidity falls, due, say, to the entrainment of dry air, this effect may become more important, while if the up-draught ceases altogether, the supersaturation would tend to establish itself in equilibrium with the whole population of droplets, and as a result the smaller ones would evaporate and the larger ones grow.

This is just the state of affairs discussed by Langmuir when he assumes that the vapour pressure is in equilibrium with drops of *average* size. His mechanism is therefore likely to be important only after the cooling of the air has ceased and not during the growth period of the cloud; however, at this stage the majority of the droplets will usually have attained a size at which the vapour pressure differential over drops of different size will be very small, and the process will proceed only very slowly.

Langmuir also neglects the influence of the nuclei and the heat balance of the droplets on their growth. A small evaporating droplet will have a lower surface temperature than a larger one in the same environment so that the

differential excess vapour pressures over their surfaces will be reduced when the heat balance term is taken into account. The presence of a hygroscopic nucleus has a similar tendency; these two effects would combine to slow down the rate of transfer of vapour to the larger drops. Both these factors are of course considered by Howell, whose work clarifies the part played by the microphysical processes in cloud droplet growth; however, the drop-size spectrum will also be influenced by the macroscopic features of the cloud development which are difficult to take into account at this stage. It would be interesting to apply Howell's technique to a mixed population of hygroscopic and non-hygroscopic nuclei together with a few giant salt nuclei, in order to estimate the role these latter types may play in condensation. The large salt nuclei found by Woodcock, having a radius of about 100 μ at 100% relative humidity, may grow at an appreciable rate by coagulation with smaller droplets, especially in clouds of high water content. The importance of the process will be pointed out in § 7.

§ 5. THE INITIAL STAGES OF ICE FORMATION—ICE-FORMING NUCLEI

In view of the generally accepted idea that the co-existence of ice crystals and water drops is *often* essential to the release of precipitation from clouds, and the fact that many clouds are composed almost entirely of ice, the study of the ice phase in the atmosphere is of fundamental importance and has received a good deal of attention in recent years.

Although appreciable supercooling of large quantities of water, e.g. lakes, ponds, is not observed, cloud droplets normally exist in the liquid state at temperatures well below 0° c. Water clouds are a common occurrence at $-10°$ c. and have on occasion been observed at temperatures as low as $-35°$ c. In the laboratory samples of a few cm^3 of highly purified water have been supercooled to $-33°$ c. while in the case of small droplets the range has been extended to about $-41°$ c.

(i) '*Freezing*' *and* '*Sublimation*' *Nuclei*

The supercooled state is metastable and its existence implies the absence of suitable nuclei to initiate freezing. The stability of supercooled clouds at temperatures above about $-20°$ c. speaks for the rarity of efficient ice-forming nuclei in the free atmosphere. Besides the freezing of liquid water there is also the possibility that ice crystals may form by the direct deposition of vapour upon solid insoluble particles, a process which meteorologists have termed 'sublimation'. The relative importance of 'freezing nuclei' and 'sublimation nuclei' in the production of ice crystals has been a controversial subject for some years. The term 'sublimation nucleus' was introduced by Wegener (1911), who suggested that microscopic particles would induce water vapour, saturated with respect to an ice surface, to crystallize in a manner analogous to its condensation into water droplets on condensation nuclei. Bergeron (1935), on the other hand, assumed that the majority of ice crystals originate from the spontaneous freezing of supercooled drops, sublimation playing only a subordinate role. Findeisen (1938) contested the view that cloud droplets would freeze spontaneously and strongly supported the sublimation-nucleus theory, attributing the stability of supercooled clouds and fogs to the absence of sublimation nuclei. He suggested that these nuclei might be particles of quartz, which because of

its hexagonal crystal habit was assumed to be isomorphous with ice. In a later paper Findeisen (Findeisen and Schulz 1944) claimed experimental evidence for the existence of two distinct kinds of sublimation nuclei (see § 5 (iii)).

If ice crystals are to grow easily on a foreign substrate by direct deposition of water vapour, the substrate should possess the same general crystal symmetry and very nearly the same lattice spacing as those of ice. X-ray analysis reveals considerable dissimilarity in the lattice spacings of ice and quartz and also that ice has hexagonal symmetry while quartz is trigonal. Moreover, Regener (1941) has shown that quartz dust, when introduced into an expansion chamber, produced no ice crystals at ice saturation but provided condensation nuclei at water saturation; similar results have been obtained by Weickmann (1942).

In view of the rather peculiar ice crystal structure it is unlikely that many substances will act in the manner visualized by Wegener. Indeed, experiments so far have shown that the great majority of nuclei fail to produce ice crystals until water saturation is reached, while no substance has been found to act *initially* at or near ice-saturation. This strongly suggests that ice crystals are rarely formed without the intervention of the liquid phase, and several recent writers, e.g. Krastanow (1940, 1941, 1944), Rau (1944) and Weickmann (1942, 1947) have therefore taken the view that most atmospheric ice crystals are formed by the freezing of supercooled water drops, freezing being initiated by a solid insoluble 'freezing' nucleus. The formation of a water *droplet* is not however a necessary stage in the production of an ice crystal; the solid nucleus may acquire a microscopic liquid film at relative humidities considerably below 100% if its surface is hygroscopic or suitable for capillary condensation. Once this film has frozen it may continue to grow by sublimation. If the liquid film is of sufficient thickness to assume the statistical characteristics of a liquid there is no point in distinguishing between it and a macroscopic droplet surrounding the solid nucleus—freezing of liquid water occurs in both cases. As the interfacial surface energy between a solid particle and the supercooled water is less than that between the solid and the vapour, it is to be expected that ice crystals will arise more easily from supercooled drops containing solid particles than from the vapour phase by direct sublimation. In order to detect whether true sublimation nuclei are present in the atmosphere, it would be necessary to expand large volumes of air at atmospheric rates and observe whether ice crystals are formed soon after *ice saturation* is reached. Experiments which so far have sampled only small volumes of air and employed high rates of expansion have failed to prove their existence.

(ii) *Laboratory Experiments on the Supercooling and Freezing of Water and Aqueous Solutions*

A study of the supercooling of small samples of water of different degrees of purity has been carried out by a number of workers. Meyer and Pfaff (1935) supercooled samples of a few cm^3 of purified water in sealed glass containers and found the maximum degree of supercooling depended on the nature of the glass. By filtering out nuclei with collodion filters of pore diameter 10^{-6} cm. they were able to supercool water to $-24°$ c. and, after repeated distillation and filtration, to $-33°$ c.

Dorsey (1938, 1948) determined the freezing points of a number of samples of widely different origins containing a large selection of natural impurities, and found that few specimens could be supercooled below about $-10°$ c. The

introduction of known impurities into highly purified water caused raising of the freezing temperature while filtration of contaminated water had the opposite effect. Dorsey believes that the freezing temperature of water containing any given foreign particulate matter is determined by the size of the particles, larger motes in general being more efficient.

Smith-Johannsen (1948) studied the freezing of small quantities of water on a dust-free cellulose-acetate membrane supported on a chilled copper plate. The water was purified by various methods including distillation, diffusion through formvar membranes and pressure filtering. Many trials showed the freezing temperature to be independent of the rate of cooling, and of the temperature to which the water was raised between freezing cycles. The average freezing temperature of the tested samples was $-19{\cdot}2°$ C., with extremes of $-18{\cdot}0°$ C. and $-20{\cdot}0°$ C. A number of powdered materials were introduced into the water in an unsuccessful attempt to make it freeze at 0° C. Aquadag (finely dispersed graphite) was the most efficient, causing freezing at $-6{\cdot}9°$ C., with silver iodide second, acting at $-8{\cdot}8°$ C. Irradiation of the water with 1 Mc/s. ultrasonic waves produced a marked depression of the freezing point (in one case to $-38{\cdot}5°$ C.) which persisted even after the sample had stood for two months in a glass bottle.

Schulz (1948) purified water until it could be supercooled to $-20°$ C., sealed it in capillary tubes together with known quantities of various foreign substances and attempted to correlate the freezing temperature with the crystal structure, solubility, surface properties and the size of the foreign particles. The most efficient freezing agent was freshly ground sea-sand operating at temperatures between $-3°$ C. and $-5{\cdot}5°$ C., while silicates in general were quite efficient. Freshly produced surfaces initiated freezing at higher temperatures than aged ones: for example, freshly ground sea-sand acted at $-3°$ C. while wind-blown sand was effective only below $-9{\cdot}1°$ C. Neither the size nor solubility of the particles appeared to be relevant factors but the presence of dissolved air caused an increase of about 2° C. in the freezing point.

Dorsey found that appreciably supercooled water could not be made to freeze by violent shaking and splashing. Schulz confirms this, adding that such attempts to provoke freezing do not succeed until cavitation occurs in the liquid. Dorsey, Smith-Johannsen and Schulz all report that while freezing generally occurs with great rapidity from many locations in the liquid, sometimes crystals can be seen to grow more slowly from widely separated nuclei and that sometimes crystallization spreads from only one centre in the whole sample.

Hollstein (1947) has investigated the supercooling of common salt solutions of various concentrations contained in small glass ampoules of about 1 cm^3 capacity. The temperatures at which ice began to separate out from these solutions were always lower than the corresponding temperatures quoted in the *International Critical Tables* (curve (*a*) of Figure 6) by a constant amount of about $-14°$ C.—the same temperature interval through which it was possible to supercool the distilled water used in making up the solutions. Hollstein's results are represented by curve (*b*) of Figure 6. The eutectic temperature ($-21{\cdot}2°$ C.) is commonly regarded as being the lowest temperature at which sodium chloride solutions can exist; below that point ice and salt crystallize out. Hollstein found however that a eutectic solution of common salt could be supercooled to $-35{\cdot}5°$ C. and a solution saturated at 0° C. to about $-40°$ C. The significance of Hollstein's results has already been discussed by Mason (1950a); if ice or salt is to separate out from a

cooling solution, nuclei must be present to act as centres of crystallization. If the solution consists of a pure salt and water free from such nuclei, it would appear possible to supercool it until ice and salt crystals arise spontaneously. Separation of ice at the temperatures shown in curve (*a*) would require the presence of freezing nuclei operative at 0° c.; such nuclei must have been absent in Hollstein's samples, where the most efficient apparently became active at −14° c.

(iii) *Laboratory Investigations on the Properties and Behaviour of Ice-Forming Nuclei*

Two main experimental techniques have been adopted: observation of the condensation products formed on a chilled metal plate with reference to its

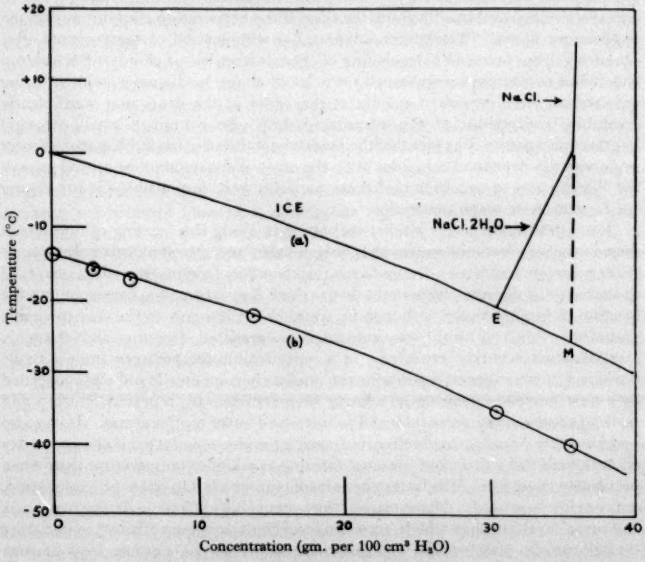


Figure 6. Phase diagram of sodium chloride solutions showing the results of Hollstein (*after* Mason). *Reproduced from the Quarterly Journal by courtesy of the Royal Meteorological Society.*

temperature, and of the sequence of water droplet and ice crystal formation in air cooled at a known rate in an expansion chamber with reference to the final temperature and humidity of the air.

The first method was used by Weickmann (1942, 1947) to determine the temperatures and humidities at which condensation products first appeared on an independently cooled nickel-plated brass foil cleaned with pure alcohol and enclosed in a cold chamber containing saturated water vapour. The metal surface could be cooled down to −40° c.; the depositions were examined with a microscope.

Measurement of the plate and air temperatures allowed the relative humidity to be calculated. In ordinary room air no ice crystals appeared on the plate at relative humidities corresponding to ice-saturation, even at temperatures as low as $-40°$ C., while at ice-supersaturations of 30% and temperatures of $-39°$ C. only a very few crystals were formed. When water-saturation was reached, however, the numbers of ice crystals increased very considerably, in consequence of which Weickmann was convinced that the majority of nuclei did not act as sublimation nuclei, but that ice particles were formed by the freezing of water drops. To investigate the nature of these 'freezing nuclei' he subjected the condensation products to a microscopic examination. When a drop was allowed to evaporate it left clearly visible residual spots which acted as centres of condensation when the relative humidity was subsequently increased to 85%. At water-saturation and temperature $-37°$ C. ice crystals appeared mainly along the edge of the evaporation ring but not on the hygroscopic spots. Weickmann advanced an explanation of these events after observing drops formed by the melting of crystals from cirrus clouds. Such a drop was found to contain several small particles of about 1μ diameter, which during evaporation were deposited mainly at the edges of the drop, and were clearly insoluble; the residue of the evaporated drop also contained small, strongly hygroscopic spots. The fact that the areas where the solid insoluble particles were preferentially deposited coincided with the areas of subsequent ice crystal growth led Weickmann to conclude that these particles were responsible for promoting ice formation at water saturation.

Rau (1944) used a very similar technique to study the freezing of droplets of diameter up to a few millimetres on highly polished surfaces of nickel or chromium. After room air had been admitted into the chamber, freezing was often observed to start at quite definite places in the drop, which Rau assumed to correspond to the location of foreign nuclei, although he could be certain only in the case of visible particles. When a drop was subjected to repeated freezing and thawing, crystallization occurred repeatedly at a reproducible temperature from a single nucleus, but after several repetitions the nucleus became sterile and was succeeded by a new nucleus operating at a lower temperature; by repeated freezing and melting supercooling was prolonged to lower and lower temperatures. It was also observed that a nucleus ineffective for freezing became reactivated if allowed to dry off at the edge of a drop and initiated freezing at a higher temperature than when previously operative. The latter phenomena, for which Rau offers no explanation, may be closely related. The surface of the responsible nucleus on drying may crack and reveal fresh surfaces which, according to Schulz, are more efficient in initiating crystallization; after several freezings and thawings the particle may become immersed in the drop and its surface become relatively amorphous and so less efficient as a freezing nucleus. Rau found that the majority of airborne nuclei caused freezing between $-10°$ C. and $-12°$ C. although a few acted at higher temperatures. The maximum supercooling he could obtain in the presence of nuclei was to $-33°$ C.

He claims, however, that when all nuclei were removed it was possible to supercool drops to $-72°$ C. and to maintain them at that temperature for several hours. Below $-72°$ C. the drops froze uniformly without the selection of special sites and generally exhibited a very fine crystal structure; with very slow cooling, however, the crystal habit could be identified as octahedral with occasional tetrahedra and cubes in contradistinction to the hexagonal crystals formed at

higher temperatures. The crystals remelted on being heated above $-72°$ C. Rau suggested that at $-72°$ C. spontaneous freezing occurred without the aid of nuclei, resulting in a new crystalline ice structure.

These latter results were questioned by Cwilong (1947 a) who could not reproduce them unless small amounts of contaminating vapours (e.g. ether, acetone) were introduced into the chamber. Rau's experimental arrangement has been criticized on the grounds that contamination could arise from the alcohol or a similar organic liquid used as the coolant, and that the electrically driven fan used to equalize the air temperature in the chamber would stir up dust and produce nuclei by arcing and wear at the commutator.

Recently Palmer and Brewer (as yet unpublished) have designed an apparatus to avoid these faults and have carefully repeated Rau's experiments; their results confirm those of Cwilong. Drops larger than about 0·2 mm. diameter supported on a polished gold or nickel surface in nucleus-free air always froze at a temperature near to $-20°$ C. Smaller drops could be supercooled to lower temperatures—those of 30 μ diameter down to $-30°$ C. The freezing temperature could not be lowered by repeated freezing and thawing of the drops. When room air containing nuclei was introduced into the enclosure the freezing temperature of the drops rose to values as high as $-5°$ C.; in this case the freezing temperature could be altered by repeated freezing and melting, but in an irregular manner rather than in steps as reported by Rau. In the presence of a small amount of alcohol the freezing temperature showed a continuous fall as the alcohol distilled over on to the drops. In the absence of airborne nuclei, water–alcohol mixtures of increasing alcohol content could be progressively supercooled to temperatures as low as $-72°$ C. The change of crystal habit with temperature in the range $0°$ C. to $-72°$ C. was identical with that observed by Rau. There seems then to be little doubt that the water drops examined by Rau were contaminated with alcohol and that so far no one has succeeded in supercooling water drops in contact with a metal or glass surface much below $-30°$ C.*

Lafargue (1950) has found, however, that droplets of water or of aqueous salt solutions suspended on spiders' threads always freeze at $-40·5 \pm 1·5°$ C. if their diameter is less than 20 μ. Larger drops freeze at higher temperatures which vary with their size and the concentration of the solute. The lower freezing temperature of smaller droplets, which appears to be an established feature in the above experiments, is perhaps due to the lower probability of their being infected by a foreign nucleus. Hollstein (1947) has made observations on droplets of concentrated common salt solutions sprayed on to the chilled metal plate of a Weickmann cooling chamber. Most of the droplets crystallized during their fall and crystals of sodium chloride were identified on the plate, but became surrounded by a water film when the relative humidity was subsequently increased. When the temperature fell below $-30°$ C. a sudden change occurred, the surface films becoming dim and rough, ice being deposited on the surface of the crystals, apparently in a crust.

The experiments on the supercooling and freezing of small samples of water and aqueous solutions which have just been described, while providing considerable information on the nature and properties of various types of nuclei, are always

* Large cubical crystals, apparently similar to those described by Rau, were grown from water–alcohol mixtures at low temperatures by Barendrecht (1896); such crystals had been supposed to be composed of a hydrate, but Barendrecht was able to obtain corresponding crystal-forms from mixtures of water with a number of alcohols and also with acetaldehyde.

complicated by the presence of a solid supporting surface or the walls of the containing vessel, which may themselves act as nucleators. It is extremely difficult to prepare solid surfaces of known ice-forming properties which will be retained for an appreciable period of time. In order to avoid the presence of such surfaces and to simulate more closely atmospheric conditions, it is very desirable to examine the condensate when suspended in air which can be cooled at a controlled rate to predetermined temperatures. The Wilson expansion chamber provides an obvious technique and has been employed by a number of workers whose results will now be summarized.

Regener (1941), having passed air through a bacteria filter, subjected it to very rapid expansions and examined the resultant condensate through a microscope. With the initial temperature below 0° c. and with expansion ratios between 1·3 and 1·5 only supercooled water droplets appeared; at 1·5 a few ice crystals appeared among the water drops and with ratios between 1·6 and 1·8 (depending on the initial temperature which is not stated) a pure ice mist was formed. From the

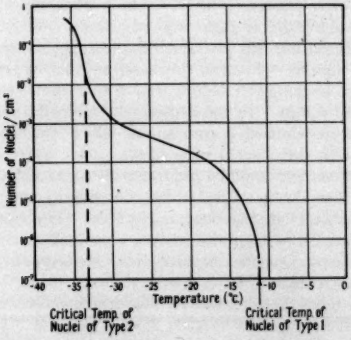


Figure 7. The average ice nucleus spectrum in city air at Prague (Findeisen and Schulz).

data given by Regener it is not possible to calculate the temperatures reached at the end of these expansions, but he reports that water drops existed at temperatures down to −50° c. This temperature, however, appears to have been calculated from the expansion ratios and the initial temperature of the air assuming a perfectly adiabatic expansion, and so may have been several degrees too low.

Findeisen and Schulz (1944) set out to study the behaviour of ice-forming nuclei in outdoor air under roughly atmospheric conditions by using a large expansion chamber (capacity 2 m³) and employing rates of expansion corresponding more nearly to those prevailing in the atmosphere. They investigated particularly the effect of varying the rate of expansion on the temperature at which ice crystals first appeared. At temperatures a few degrees below zero, water drops alone occurred; in the temperature range −6° c. to −20° c. ice particles appeared in concentrations of only about 1/m³, but their number gradually increased as the temperature was lowered. At −32° c., however, a sharp increase in the number of ice particles occurred, the concentration rising from about 1/litre to about 1/cm³ (see Figure 7). At still lower temperatures the numbers increased still further but

even at −40° c. (the lowest temperature attained) the concentration of ice particles was considerably less than that of water drops. That some ice particles could be formed before water saturation was reached was established by interrupting an expansion before the condensation point was reached; high ice-supersaturations were produced but droplet formation was prevented. The number of ice crystals appearing under such conditions was said to be not noticeably less than when crystals were produced in mixed clouds (Findeisen 1942). Findeisen was led to the conclusion that there were two distinct kinds of *sublimation* nuclei, those of the first kind being characterized by their rarity and high efficiency for ice formation, and those of the second kind by their relative abundance but ability to act only below about −32° c. By varying the rates of expansion Findeisen was able to produce a range of equivalent vertical speeds of from 5 to 22 m/sec. The critical temperature at which the 'Type 1' nuclei became active fell when the rate of expansion was increased, showing that they were less effective when subjected to rapidly increasing relative humidities. It was also established that the dependence

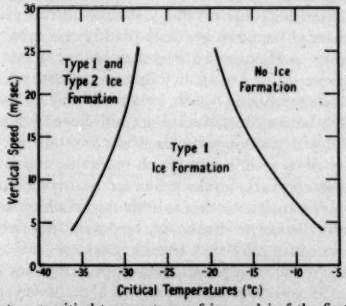


Figure 8. Relation between critical temperatures of ice nuclei of the first and second types and the rate of expansion (expressed as a rate of vertical ascent in the atmosphere) during the initial condensation (Findeisen and Schulz).

was upon the rate of expansion which existed at the moment of cloud formation when no ice particles had yet formed, the critical temperature being independent of the vertical speed prevailing after the commencement of drop formation. The 'Type 2' nuclei, on the other hand, initiated ice formation at higher temperatures when the rate of expansion was increased. These features are shown in Figure 8.

Cwilong (1947 b, c) examined outdoor air using an expansion chamber of only a few hundred cm^3 capacity and high rates of expansion. He observed no ice crystals above −32° c., probably because of their rarity and the small volume of his chamber. At −33° c. a few ice particles appeared among a much larger number of water drops; at lower temperatures their concentration increased but became comparable to that of the water drops only when the temperature fell below −41° c. Measurements on the Jungfraujoch and over mid-ocean indicated that the −32° c. nuclei were absent or extremely rare, no ice crystals appearing in the chamber until a temperature of −41·2° c was reached. Later, Cwilong (1949) observed that the concentration of the −32° c. nuclei relative to that of the −41° c. nuclei at Wellington (N.Z.) was lowered by the onset of a sea breeze. These three observations all suggest that the −32° c. nuclei may be of land origin.

Fournier d'Albe (1949) used a somewhat larger chamber and illuminated a very small volume to facilitate the microscopic observation of individual particles. With outdoor air he obtained results in agreement with Cwilong. In the temperature range 0° c. to −32° c. only water droplets appeared. At about −33 °c. a few ice crystals appeared simultaneously with the water drops (i.e. at the dew point) and did not increase in number after the expansion was completed; the critical temperature was not definite but varied between −32° c. and −36° c. When the temperature fell below −41° c. a sudden increase in the number of ice crystals was observed, the fog being composed almost entirely of ice. Again, these crystals did not appear at the frost point (ice saturation) but only when the dew point (water saturation) was reached. Using carefully cleaned air Fournier d'Albe found that with relative humidities of 112% relative to water (188% relative to ice) neither droplets nor crystals were formed even at temperatures below −42° c. An interesting aspect of his work was the introduction of artificial nuclei of various pure salts into an expansion chamber containing clean air and the determination of the temperature and humidity at which ice particles appeared. Nuclei of sodium chloride (cubic crystals), sodium nitrate (hexagonal), caesium iodide (cubic) and those obtained by heating a platinum wire to redness, all producd ice crystals only below −41° c. and at the dew point. Crystals of silver iodide produced ice crystals at −7° c. but again only at water-saturation and so appeared to be particularly efficient freezing nuclei, presumably by virtue of their hexagonal symmetry and of their lattice constants being quite close to those of ice. Peculiar results were obtained with cadmium iodide whose crystal structure is not quite so close to that of ice as silver iodide, but which is soluble and slightly hygroscopic. During the first expansion CdI_2 nuclei produced ice crystals only below −41° c. and at the dew point as did all the other soluble nuclei which were tested. When the ice particles were allowed to evaporate, however, they left behind nuclei on which ice would grow again after very small expansions, producing temperatures only very slightly below the frost point—they appeared to act as true sublimation nuclei. If these nuclei were warmed above −9·1° c., however, they ceased to act in this manner, subsequent large expansions producing only water drops. Mason (1950 a) has suggested that warming to temperatures below −9·1° c. causes only partial evaporation of the ice particle, leaving a microscopic oriented film of ice on the surface of the crystals which can be deformed into exact fit with the substrate. At temperatures above −9·1° c. the film probably assumes liquid characteristics, a view supported by the fact that the critical temperature of −9·1° c. is very close to the eutectic temperature of CdI_2 solution.

Palmer (1949) has used Cwilong's cloud chamber on an aircraft and reports that ice nuclei which become active at −32° c. in the air of surface layers were absent above inversions, stratus cloud or haze-tops, but that the −41° c. nuclei exist at all heights up to 6 km., the ceiling of the observations. He has also used a 5-litre chamber in the laboratory in order to detect the Findeisen ' Type 1 ' nuclei which are active above −32° c. Using expansion rates corresponding to vertical speeds of 100 to 200 m/sec. he found nuclei which produced ice crystals between −18°c. and −31°c., their numbers increasing about 500-fold over this temperature range. In common with Findeisen, Palmer has observed occasional ice crystals at high ice supersaturations, but without a water cloud being formed.

Schaefer (1946, 1948) has used a very simple apparatus to study the production of ice crystals in a cloud of supercooled water droplets. It consists essentially of a

commercial freezing unit of about 70 litres internal capacity in which the air temperature is about $-20°$ C. at the bottom; with the top open and at room temperatuae the inversion makes the air in the chamber very stable. A supercooled cloud is produced either by breathing into the chamber or by evaporating water from a damp cloth. The cloud is illuminated by a powerful parallel beam of light, the walls of the chamber being lined with black velvet to provide dark background illumination. Ice particles suspended in the cloud glitter in the light beam and their concentration can be estimated. They soon grow at the expense of the water drops and fall to the bottom of the chamber where they may be caught on slides for subsequent microscopic examination, or in a tray of polyvinyl formal dissolved in ethylene dichloride, which permits exact plastic replicas of the crystals to be made. This apparatus has been used by Schaefer (1950) to study the efficiency of various substances as ice-forming nuclei; the substance was introduced into the cold chamber in finely dispersed form, the temperature at the bottom of the chamber when ice crystals first appeared being noted. Silver iodide was the most efficient nucleator, acting between $-4°$ C. and $-10°$ C., while most kinds of soil particles produced ice crystals between $-12°$ C. and $-22°$ C. The same technique has been used by Schaefer (1949 a, 1950) to measure the concentration of ice-forming nuclei in the atmosphere on Mt. Washington. Concentrations from $1/m^3$ to $10^7/m^3$ have been found, with 55% of the 4,500 observations giving counts of less than 10^4 per m^3. Schaefer (1946, 1948) also discovered that a supercooled water cloud may be converted into ice by chilling it below $-39°$ C. This may be accomplished by dropping into the cloud a fragment of solid carbon dioxide (dry ice); Mason finds that a pin-head fragment will completely transform 10 litres of supercooled cloud producing some 10^7 crystals, specimens of which are shown in Plate 12. Schaefer measured the critical temperature at which these large numbers of ice crystals appeared by dropping a pellet of solid mercury (melting point $-38{\cdot}9°$ C.) into the cloud. It left large numbers of ice crystals in its wake as long as it remained solid, but at the instant of melting this property disappeared.

Vonnegut (1949 b) has made a detailed investigation of the various methods of producing silver iodide smokes and of the effects produced by introducing them into a supercooled cloud in a cold chamber. The smoke is most conveniently produced by burning a solution of AgI in acetone in a hydrogen flame, when particles ranging from about 30 A. to 1,400 A. in diameter are produced. Some of these nuclei may cause freezing at temperatures as high as $-4°$ C. but the majority become effective at about $-10°$ C. The ice crystals so produced are mainly in the form of small hexagonal plates.

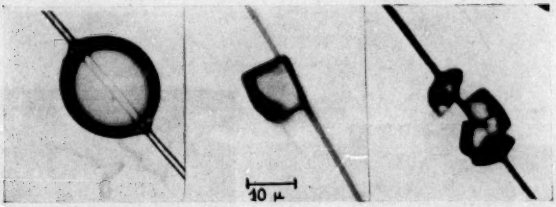
(iv) *Discussion of Experimental Results*

The results of the foregoing experiments have recently been examined by Mason (1950 a) with a view to deducing the nature of the ice-forming nuclei in the atmosphere. The experimental evidence points to the existence of three main types of nuclei characterized by the temperature ranges over which they are effective, viz. $0°$ C. to $-32°$ C., $-32°$ C. to $-41°$ C. and just below $-41°$ C. The first two groups are to be identified with the 'Type 1' and 'Type 2' nuclei of Findeisen.

It appears that the nuclei which produce ice crystals between $0°$ C. and $-32°$ C. become effective only at or near water-saturation, which suggests that they

initiate freezing of the liquid water accumulated on their surfaces rather than allow crystal growth by direct deposition of the vapour, and so they may be described as 'freezing nuclei'. Experiments so far have failed to establish that true 'sublimation nuclei' (acting at ice-saturation) exist in the atmosphere. The appearance of small numbers of ice crystals at high ice-supersaturations (Weickmann, Findeisen, Palmer) may be a consequence of the nuclei acquiring liquid films by capillary condensation at humidities well below water-saturation. The observations of Rau, Schulz and Dorsey that, in this temperature range, ice formation starts at definite spots in the liquid suggest that the responsible nuclei are solid and insoluble, a conclusion supported by Meyer and Pfaff, who were able to remove them by filtration and so to supercool water to $-33°$ C. One would expect the efficiency of a freezing nucleus to depend mainly on its crystalline and surface structure; in general, its size will not be of primary importance although a larger nucleus will have a higher probability of possessing a suitable site for ice-crystal formation—oriented overgrowth tends to occur preferentially along steps or cracks in the surface where there are unsaturated molecular forces. Findeisen's observation that the critical temperature for ice crystal formation was lowered when the rate of expansion *prevailing at the moment of condensation* was increased suggests that the efficiency of an insoluble particle as an ice nucleus depends on whether it is first involved in droplet condensation. Higher rates of expansion produce greater peak supersaturations and hence are more likely to allow droplet formation on insoluble particles. Why these particles should initiate crystallization less readily when immersed in a droplet than when surrounded only by water film (as will be the case for low rates of expansion) is not clear, but is compatible with Rau's report on the sterilization of a nucleus by immersion in bulk liquid and its reactivation by the partial drying of its surfaces.

The experiments of Meyer and Pfaff indicate that solid insoluble nuclei of diameter greater than 10^{-6} cm. cause water to freeze above $-33°$ C., while samples containing natural nuclei generally freeze well above this temperature. Hence the thousand-fold increase in the number of ice particles occurring in expansion chamber experiments at this same temperature suggests the introduction of a new type of nucleus—a view supported by the fact that the nuclei operating between $-32°$ C. and $-41°$ C. react to changes in expansion rate in the reverse manner to the nuclei acting at higher temperatures. Findeisen's suggestion of a second type of *sublimation* nucleus becoming operative at about $-32°$ C. is disproved by Fournier d'Albe's observations that these nuclei are effective only below the dew-point, which again suggests that freezing of liquid water is involved, Hollstein's work on the freezing of common salt solutions, and her observations that ice forms on sodium chloride crystals below about $-30°$ C., together with the fact that sea salt particles may well be present in sufficient numbers, suggest that the new nuclei may consist of sea salt. Below about 60% relative humidity they will tend to exist as crystals and at higher humidities as droplets of solution. In the expansion chamber experiments they will consist of very dilute solution, which, if composed of chemically pure salt and water free from foreign particles, will freeze only below $-41°$ C., as shown by Fournier d'Albe and Lafargue. In the presence of foreign particles, however, ice will begin to separate out at a temperature determined by their efficiency as freezing nuclei, but complete solidification will not occur until the 'effective' eutectic temperature is reached, which, in Hollstein's experiments was about $-35°$ C. This temperature will vary with the purity of the salt, which provides a possible explanation for the range of critical temperatures



(*a*) (*b*) (*c*)

PLATE 1.
Large droplet-nucleus from haze, (*a*) before and (*b*), (*c*) after desiccation (Dessens).
(*Reproduced from the Quarterly Journal by courtesy of the Royal Meteorological Society*)

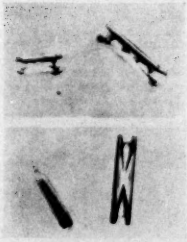


PLATE 2.
Single hollow prisms of isolated cirrus (−50° C.).

PLATE 3.
Clusters of hollow prisms (−44°C.).

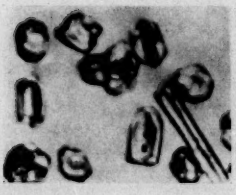


PLATE 4.
Complete prism and prism with pyramidal ends. Cirrostratus (−47° C.).

PLATE 5.
Complete prisms. Cirrostratus (−26° C.).

To face page 176

PLATE 6.
Thin hexagonal plates. Nimbostratus (−13° C.).

PLATE 7.
Richly branched snow star (advanced dendritic growth).
(*Reproduced from 'Snow Crystals' by Bentley and Humphreys by courtesy of the McGraw-Hill Book Co. Ltd.*)

PLATE 8.
Crystal showing successive transitions from plate to stellar growth.
(*Reproduced from 'Snow Crystals' by Bentley and Humphreys by courtesy of the McGraw-Hill Book Co. Ltd.*)

PLATE 9.
Clusters of ice needles. (*After* Nakaya)

PLATE 10.
Snowflakes composed of agglomerations of stellar crystals. (*After* Findeisen)

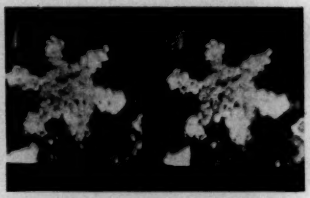


PLATE 11.
Stellar crystal covered by frozen droplets.

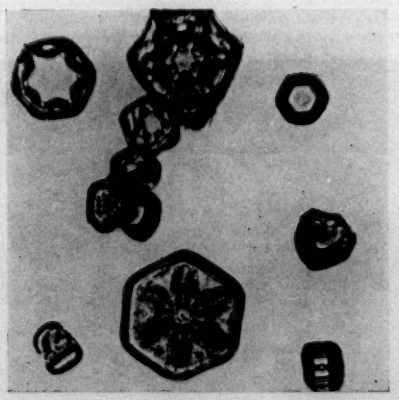


PLATE 12.

Plates 2, 3, 4, 5, 6, 11 *are reproduced from Weickmann's monograph* '*Die Eisphase in der Atmosphäre*' (1947).

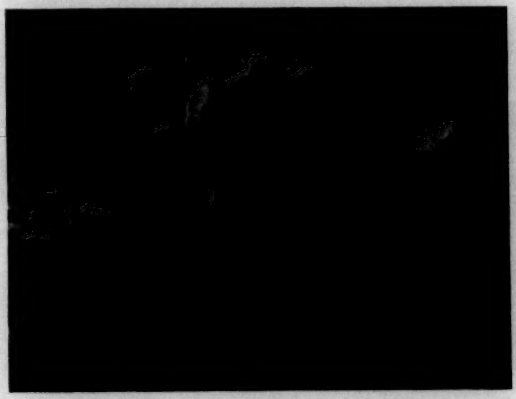


PLATE 13.

Cumulus and cumulonimbus (shower-cloud). In the foreground is a towering cumulus cloud with characteristically well-defined edges. The larger cloud in the background has the ragged and fibrous edges which develop following the glaciation of the summits of a cumulus, and which are associated with its transformation into a shower-cloud.

PLATE 14.

found by Fournier d'Albe. We still have to explain why these nuclei, assumed to consist of droplets of sea salt solution containing solid particles, produce ice crystals at higher temperatures when the rate of expansion is increased. At high rates of expansion, and so high rates of cooling, the salt may not be able to diffuse quickly enough to keep pace with the increasing growth of the droplet, thus causing a concentration gradient to be set up. The outer layers of the drop will therefore in effect consist of pure water, which in the presence of a suitable insoluble nucleus will freeze at a higher temperature than if salt were present; solidification of the interior may not be completed until its temperature falls below the 'eutectic' temperature. The above assumptions afford an explanation of the observations that the $-32°$ C. nuclei are relatively less abundant over the sea, at high levels, on the coast with a sea breeze blowing, and above inversions or haze-tops. In these localities and under these conditions, the salt nuclei will have a smaller chance of becoming contaminated by foreign particles from land sources, and so will be more likely to operate below $-41°$ C. than above.

It appears that droplets of pure water or solutions of pure salts freeze at a definite temperature of just below $-41°$ C., this critical temperature apparently being independent of the nature of the salt. This appears then to be the temperature at which water freezes spontaneously without the aid of foreign nuclei; there seems to be no conclusive evidence that water can be kept in the liquid state at temperatures below $-41°$ C.* It seems very probable that the formation of large numbers of ice crystals by chilling a supercooled water cloud to below $-39°$ C., as described by Schaefer, was the result of freezing of very small water drops formed spontaneously from the vapour. The discrepancy between this value and the $-41°$ C. obtained from expansion chamber experiments may be partly due to the fact that Cwilong and his co-workers were dealing with rapidly growing drops whose surface temperature would be higher than the air temperature, and partly due to the difficulty of accurately deducing final temperatures at the end of a rapid expansion.

That Fournier d'Albe could produce no condensation products on cooling clean saturated air below $-41°$ C. is explained by the fact that he employed only very small expansion ratios and so produced only small supersaturations; Schaefer, on the other hand, by introducing a pellet of solid mercury into a supercooled cloud at $-10°$ C. produced supersaturations of the order 1,000%.

§6. THE GROWTH OF ICE PARTICLES

The ice elements formed in natural clouds are of three main types: individual ice crystals (or groups of crystals having a common nucleus), snowflakes and hailstones. The ice crystals grow in an ice-supersaturated atmosphere by the diffusion of water vapour to the growing surfaces and may exist as individual units of simple geometrical shape (e.g. hexagonal prisms and plates), or under suitable conditions may grow rapidly as dendritic forms into richly branching snow crystals. Several of the latter crystals may collide and coagulate to form a snowflake. A hailstone, on the other hand, may originate from a simple ice crystal or from a frozen droplet, but its subsequent growth results predominantly from collision with supercooled water droplets. The factors which decide whether ice crystals and snowflakes or hailstones are produced are discussed in §7. Snowflakes may also collide with supercooled drops which freeze on contact to produce a form of granular snow (Plate 11).

* There is some evidence, from reports of aircraft icing (e.g. Aanensen 1948) and from optical phenomena in very high clouds (e.g. Störmer 1948), which suggests that droplets may occur at temperatures considerably below $-41°$ C.

(i) *The Formation of Ice Crystals and Snowflakes*

The collection, examination and classification of ice crystals has been carried out by a large number of workers, the observations having revealed a wide variety of shapes and forms (see *Snow Crystals* by Bentley and Humphreys (1931)). Little progress has been made however in explaining their diversity and how their external morphology is related to the crystalline structure and to the conditions of growth. A number of early attempts were made to correlate the frequency of various crystal shapes with the prevailing temperature, but Wegener (1920) was the first to realize that the ice-supersaturation of the environment was an important factor. He suggested that at slight ice-supersaturations complete crystals were formed, while high supersaturations favoured skeletal growth. Thus, he interpreted the crystals giving rise to haloes in cirro-stratus cloud as being complete, while those existing in cirrus without halo phenomena were assumed to be skeletons, a view supported by the recent observations of Weickmann (1947). Wegener believed, however, that every crystal started life as a skeleton form, growing into a complete crystal only when subsequently subjected to low supersaturations. More recently, similar ideas have been advanced by Findeisen (1938) but examination of crystals reveals that the reverse is more often the case (Mason 1951). Findeisen (1938) and Krastanow (1940) among others, assert that skeletal or dendritic crystals forms are the consequence of high ice-supersaturations, while complete crystals of simple geometrical shape are produced at low supersaturations. Observations show however that the most richly branching snow stars are confined to low level clouds of relatively high temperature and liquid water content but of relatively low ice-supersaturation (see Table 5). The earlier attempts to correlate the predominant crystal habit with the meteorological conditions on the ground at the place of observation were inconclusive and often contradictory, which perhaps is not surprising since only the conditions prevailing during growth of the crystal are relevant. Such information has only recently become available as a result of the work of Weickmann (1947) during World War II. He has made a systematic study of ice crystals and snowflakes both at ground levels, and at all heights in the atmosphere up to cirrus levels, in all types of clouds, and consequently in widely different conditions of temperature, ice-supersaturation and vapour density, the temperature of the sampling region being measured. The ice crystals were caught on a glass slide coated with a lacquer, which after a short exposure to the cloud elements was brought into the aircraft and photographed. His results are summarized in Table 5.

Krastanow (1943), Weickmann (1947) and Wall (1947) have all made attempts to interpret their observations in a qualitative manner in terms of the Becker and Döring–Volmer (Volmer 1939 b) theory of crystal growth by surface nucleation. This mechanism, as we shall see, is much too slow to account for the observed rates of crystal growth and appears to be of little value is discussing the development of macroscopic crystals. A recent attempt has been made by Mason (1950 b, 1951) to explain the development of the various crystal forms in terms of their molecular and surface structure, their fall motion and the meteorological conditions prevailing in the parent cloud. A short account of this work will now be given.

From the molecular structure of ice as revealed by x-ray and neutron diffraction it is concluded that it is profitable to speak only of a mean or average crystalline structure which is either dihexagonal bipyramidal (D_{6h}^4) or ditrigonal bipyramidal (D_{3h}^4), both space groups being non-polar. This is in disagreement with the views

Table 5. Classification of Ice Crystals according to Shape, Cloud Type and Environmental Conditions (Based on Weickmann's Observations)

Crystal form	Cloud type	Conditions in sampling region: Temperature	Humidity* (w.r.t. ice)	Vapour pressure*
Hollow prisms (Mainly clusters)	Cirrocumulus, Cirrus Castellatus, Cirrus Uncinus	−30° C. to −50° C.	135 to 165%	0·50 to 0·06 mb.
Complete prisms	Cirrostratus	−25° C. to −50° C.	75 to 100% − W; 95 to 165% − I	0·60 to 0·06 mb.
	Altostratus, Altocumulus	−25° C. to −50° C.	125 to 165%	0·80 to 0·06 mb.
Thick plates	Altocumulus, Altostratus	−15° C. to −25° C.	115 to 125%	1·9 to 0·80 mb.
Thin plates	Nimbostratus, Stratus, Stratocumulus	0° C. to −15° C.	100 to 115%	6·0 to 1·9 mb.
Stars, Needles (Dendritic forms)	Nimbostratus, Stratus, Stratocumulus	0° C. to −15° C.	100 to 115%	6·0 to 1·9 mb.

* Calculated on the assumption that water saturation exists in all clouds except cirrostratus.

of Weickmann (1947), who assumed ice to belong to the dihexagonal-pyramidal space group, C_{6v}^4, which is polar and should therefore exhibit pyro-electricity. Although all experimental attempts to detect pyro-electricity in ice have failed, Wall (1947) and Rossmann (1948) suggest that it plays an important part in the production of thunderstorm electricity.

The dominant growth forms to be expected in the D_{6h}^4 and D_{3h}^4 crystal classes are hexagonal prisms, hexagonal bipyramids and hexagonal prisms with pyramidal ends. The external morphology of a crystal is however determined by the relative rates of growth of the different crystallographic faces; the pyramidal faces, being of lower atomic density (i.e. of higher index), will in general grow faster than the basal and prism faces and so tend to grow out. On these grounds the dominant crystal habit of an ice crystal should be a hexagonal prism bounded by the slow-growing (0001) and $(10\bar{1}0)$ or $(11\bar{2}0)$ faces—this is certainly the most common form in the atmosphere. The pyramidal faces may persist for some time when the crystal is growing or evaporating very slowly. This is most likely to occur in cirrostratus where a very slow and steady lifting process may continue for several hours, or the cloud may decay equally slowly. It can be shown that the sides of the pyramids should be inclined at 28° to the longitudinal axis of the crystal. Humphreys (1929) shows in fact that pyramidal forms are necessary to explain the formation of some haloes seen occasionally in cirro-stratus, and Weickmann (1947) obtains good agreement between calculated and observed values of the angular diameter of their rings by assuming the inclination of the pyramidal faces to be 27° 57′.

In the examination of the growth-rate of an ice crystal, it is shown that, according to the Becker–Döring theory of growth by two-dimensional nucleation on a perfectly flat crystal surface, a maximum linear growth-rate of only 10^{-32} cm. sec^{-1} will be attained under the highest ice-supersaturation prevailing in natural clouds; this is many orders of magnitude too small. On this theory the growth rate is largely controlled by the rate of appearance of two-dimensional nuclei formed on the surface by the chance aggregation of molecules. If however the surface were to contain a number of molecular steps, growth could ensue by the deposition of molecules at the edges of the steps, which would thus themselves act as two-dimensional nuclei. Recently Frank (1949) has suggested that such steps can arise and persist during growth by the termination of 'screw dislocations' in the crystal surface. The theory has been worked out mathematically by Burton and Cabrera (1949) and Burton, Cabrera and Frank (1951) and applied by Mason to the initial stages of ice crystal growth in a water-saturated environment. It is shown that the growth rate attains a maximum value of 35μ sec^{-1} at about $-13°$ c., a much more realistic value than that predicted by the Becker–Döring theory. The growth rate as calculated on this theory applies only during the very early stages when the crystal dimensions are less than $0{\cdot}1\mu$; thereafter the growth rate is controlled by the diffusion of water vapour to, and the conduction of the latent heat of crystallization away from, the crystal surface. The above theory gives the *maximum possible* rate of growth; the average rate of growth when diffusion rules is considerably less than this maximum value, the radius r (cm.) of a stationary spherical crystal after a time t (sec.) being given approximately by the equation $r = 3{\cdot}5 \times 10^{-4} \sqrt{t}$ when the temperature is $-13°$ c. In the atmosphere, then, the rate of growth of a crystal will be determined by the rate of arrival of vapour at its surfaces and the rate at which the molecules can be assimilated into the surface structure. The relative rates of these two processes will determine the mode as well as the rate of growth, as will be seen later.

In discussing the growth forms of atmospheric ice crystals and their marked dependence on temperature, it is pointed out that the hexagonal prisms which predominate below $-25°$ C. (Plates 2–5) develop as a result of preferential growth along the principal (c) axis, i.e. normal to the basal (0001) face. It is suggested that this is a consequence of molecules on the basal face requiring smaller activation energies for surface diffusion than those on the prism faces and of the atomic binding being weaker in directions normal to the basal plane, as indicated by ice cleaving preferentially parallel to this plane, by the coefficient of linear expansion being greater along the c axis than along the a axis, and by the basal plane giving a very intense diffuse x-ray reflection. Prismatic columns appear both singly and in clusters. Clusters are very common in isolated cirrus clouds, the crystals containing very marked cavities as shown in Plate 3. They are entirely absent however in cirrostratus. The crystals constituting a cluster appear to originate from a common nucleus, in all probability a frozen droplet. It is suggested that when the droplets freeze dendrites will shoot through the drop and terminate in the surface to form suitable sites for the origin of the prismatic crystals. If the initial nucleus is not a droplet but a solid particle possessing only a microscopic water film (as will almost certainly be the case in cirrostratus), simple crystals will tend to form. Cluster shapes occur relatively less frequently in the lower supercooled water clouds, probably because the great majority of ice nuclei are produced by the splintering of existing snow crystals, although the small clusters of needle-like crystals (Plate 9) which are sometimes precipitated from ' warm ' clouds are very similar to those found in isolated cirrus and probably arise in much the same way.

Weickmann found a continuous transition from long, thin prisms at about $-40°$ C. through shorter, thicker prisms to thin hexagonal plates, which predominated above $-10°$ C. It appears that the thin plates are formed when growth normal to the basal surfaces is overshadowed by a much more rapid tangential development along the secondary axes (Plate 6). To explain this transition from columnar to plate forms, Weickmann suggested that at higher temperatures the molecules striking the basal faces do not form a new layer but diffuse across the face and are built into the crystal structure only at the edges. Such migration of atoms from one crystal face to another has been observed experimentally (e.g. Volmer and Estermann 1921). Mason has estimated quantitatively the importance of this effect, and shows that in the most favourable case the relative rate of growth of the edges of the plate crystal by surface diffusion from the basal faces and by direct deposition of the vapour is given by $2\lambda_s/t$, where λ_s is the distance a molecule will travel on the crystal surface before evaporating and t is the thickness of the plate. The magnitude of λ_s is calculated and it is shown that for surface diffusion of molecules to be the dominant feature, t must be less than 50μ. Now the thin ice plates found in the temperature range $-20°$ C. to $0°$C. have thicknesses of the order 10μ to 20μ so that migration of molecules from the basal faces to the edges may be a significant factor in their development. The surface diffusion of molecules will be reduced at lower temperatures, as indicated by the formation of ' glassy ' ice at temperatures below about $-70°$ C. There is another factor which will assist in the plate-like development. When the Reynolds number of the falling plate exeeds about 0·1, corresponding to a plate diameter of about 100μ, it will fall with its principal axis vertical; the concentration gradient in the vapour will then be greatest towards the edges of the falling plate, leading to enhanced tangential growth of the crystal.

Of the dendritic growth forms, the snow stars are the most striking (Plate 7). They generally arise from small hexagonal plates, the corners of which shoot out to form the arms of the star. Such a development marks the beginning of dendritic growth as opposed to the ordered development of crystal faces leading to polyhedral growth. Polyhedral growth occurs if the diffusion field is not disturbed appreciably from its equilibrium state. If, however, the crystal grows rapidly the field is continuously disturbed from its equilibrium state as the crystal pushes its way through it. The field will strive continuously towards its equilibrium pattern with a speed determined by the coefficient of diffusion, but if the growth rate is high, non-uniform concentration gradients will be set up and the rate of arrival of material at the crystal will be slowed down. If the material is extracted from the field at a higher rate than it can be replaced by diffusion the concentration in the immediate neighbourhood of the crystal will fall below the equilibrium value and only the corners will grow, since they protrude into regions of higher concentration than exists near the rest of the crystal surface. Polyhedral growth thus breaks down and dendritic growth begins. The diffusion field will reorganize itself and become highly convergent towards the growing tips which will continue to penetrate into new regions of high concentration. The criterion for this transition from polyhedral to dendritic growth must express the relative tendencies for the diffusion field to be disturbed by the growing crystal, and for it to return to its equilibrium condition. For a crystal at rest in the vapour Mason shows the criterion to be Rd/D, where R is the linear growth rate, d the diameter of the crystal, and D the coefficient of diffusion. If however the crystal is falling, the diffusion field will be disturbed at a rate more nearly equal to the terminal velocity v of the crystal than to the growth velocity R. In this case the criterion becomes vd/D. If $vd/D \gg 1$, the field will be greatly perturbed from its equilibrium pattern and dendritic growth will be favoured. Conversely, if $vd/D \ll 1$, the disturbance will tend to become smoothed out and polyhedral growth will persist. The terminal velocity of the thin hexagonal plate crystals is about 30 cm. sec^{-1} in still air, and $D \sim 0{\cdot}3$ $cm^2 sec^{-1}$; thus dendritic growth will set in only when d considerably exceeds about 0·1 mm. This is in good agreement with observation. Whether or not this dendritic development will be well marked will depend also on the vapour density in the cloud; richly branching forms will occur only in clouds of high vapour concentration. A single crystal may show more than one well-marked transition from stellar to plate-like development. In Plate 8 the spaces between the arms of a star appear to have become filled in during subsequent growth to form a large plate; dendritic growth was resumed at a later stage. Spatial variations in supersaturation, vapour density and vertical velocity in the cloud are to be expected, the above transitions being imposed by changes of environment in the fall path of the crystal.

Snowflakes consist of a rather haphazard agglomeration of loosely interlocking crystals, the components being mainly stellar crystals and needles (Plate 10). The individual components can often be identified and separated, and it can be inferred that the flake is built up as a result of collision and subsequent coagulation of individual elements of different terminal velocities. In addition snow crystals generally possess appreciable relative horizontal velocities which may be important in the collision process as indicated by the formation of very large flakes apparently by the sideways contact of stellar crystals. Coagulation is greatly facilitated if the surfaces of the crystals are wet. Large snowflakes are thus to be expected when

dendritic crystals are formed in 'warm clouds' of high vapour density (temperature above about $-15°$ C.) and fall into still warmer regions where coagulation occurs. In 'cold' clouds of low water content individual crystals will predominate.

(ii) *The Formation and Growth of Hail*

The growth of an ice crystal may be greatly accelerated within a supercooled cloud by collision with droplets, which become frozen upon its surface. In clouds of small droplet-concentrations (water content considerably less than 1 gm/m^3) dendritic growth of the crystal may begin before this coagulation process adds materially to the rate at which ice is added direct from the vapour, so that the fall speed of the crystal remains of the order of 0·5 m/sec. and the coagulation with droplets is always of minor importance. This condition is also quickly reached in clouds of high liquid water content when ice crystals are introduced in numbers comparable with those of the droplets, as may occasionally occur at low temperatures. If, however, as is usually the case, the ice crystals are relatively rare, and there is an unlimited supply of droplets for the growth of each, then in a dense cloud the rate at which a crystal gains mass by coagulation becomes equal to the rate of deposition from the vapour some time before the radius of the crystal reaches about 100 μ, as may be seen from the following simple estimation.

The growth of the ice crystal, considered as a sphere, is described by the equations

$$\frac{dm}{dt} = 4\pi r^2 \rho \frac{dr}{dt} = \underset{\text{(coagulation)}}{E\pi r^2 v W} + \underset{\text{(sublimation)}}{C 4\pi r D \Delta p}, \qquad \ldots\ldots(11)$$

where dm/dt is the rate of gain of mass, ρ is the density of the added material, v is the fall speed relative to the cloud droplets, W is the concentration of water in the form of droplets, D is the coefficient of diffusion of water vapour in air, and Δp is the difference between the vapour concentration in the environment and that over the crystal surface. E is the fraction of water droplets in the path of the ice crystals with which collision actually occurs, to be estimated in the way described by Langmuir (1948), while C is a correction coefficient to take account of the motion of the crystal through the air, determined by Frössling (1938).

When the coagulation and diffusion processes are of equal importance

$$E\pi r^2 v W = C 4\pi r D \Delta p \text{ or } E r v W = 4 C D \Delta p. \qquad \ldots\ldots(12)$$

Now in the interior (at the $-15°$ c. level) of a cumulus cloud of temperate latitudes, $W \sim 4 \times 10^{-6}$ gm/cm^3, $D \sim 0{\cdot}33$ cm^2 sec^{-1}, and $\Delta p = 0{\cdot}22 \times 10^{-6}$ gm/cm^3 and, assuming that Stokes' law is obeyed, $v \sim 10^6 r^2$ cm/sec.

Inserting these values, equation (12) becomes

$$10^6 E r^3 = 0{\cdot}07 C \quad \text{or} \quad r^3 = 70(C/E) \times 10^{-9}.$$

This equation is satisfied for $r \sim 60\,\mu$, when the corresponding values of C and E are about 1·6 and 0·5 respectively.

Beyond this stage the rate of coagulation, being proportional to a higher power of the radius, soon greatly exceeds the rate of deposition of vapour, and indeed under the conditions described above, this latter process becomes negligible when the radius exceeds about 200 μ. The ice particle now grows, perhaps to a radius of a centimetre or more, as a rather compact, roughly spherical aggregate which contains air enclosures, appearing white and opaque. Its density may be as low as 0·3, and the material may be sufficiently fragile for the ice pellet to break on striking the

ground; it is called ' soft hail ', in distinction from true hail, which possesses an outer coat, and sometimes inner layers, of clear, hard ice. The fall-speeds of hailstones amount to at least several metres per second, and their production requires the presence in the parent cloud of upward air-currents with speeds of that order; it was once thought that very large hailstones whose terminal velocity must exceed 50 m/sec. indicated up-currents of the same speed, but Schumann (1938) showed that in deep supercooled clouds, speeds of about 10 m/sec. are sufficient to allow their growth, and these values commonly attained (see, for example, Byers and Braham 1948) within vigorous convective clouds, whose interiors also contain the required high liquid water concentrations.

The rate of growth of ice particles in such clouds has been estimated by Ludlam (unpublished), and it appears that a hailstone of radius 1 cm. can be produced within about ten minutes of the initial crystal forming upon its nucleus. During this period the growing hailstone may be lifted several thousands of metres before it gains a fall-speed which allows it to fall through an up-draught of 10 m/sec. As the hailstone then descends towards the 0° c. level a stage is reached when not all of the supercooled water picked up can be frozen, even though the air temperature may still be as low as −20° or −30° c. (Ludlam 1950). At this stage the surface temperature of the hailstone has risen to 0° c., and the rate at which collected supercooled water can be frozen is controlled by the rate at which the liberated heat of fusion can be disposed of to the environment. It may be shown that this condition severely restricts the growth of a large hailstone, which may pick up as much as 20–30 cm^3 of liquid water which cannot be frozen during its fall to the 0°c. level, where of course its growth entirely ceases—below this level it begins to melt. Much of this excess water must be shed as large drops into the wake of the hailstone; these drops then also grow quickly by coagulation with cloud droplets, and so in this way a few hailstones can initiate the production of rain in supercooled cloud, as is further discussed in § 7. Growth during the period when the hailstone has a liquid skin proceeds by the deposition of clear ice containing no air enclosures. Examination of large hailstones (radius exceeding about 1 cm.) often shows them to have several concentric layers of alternately clear and opaque ice about an opaque core, so that the critical condition for the appearance of a wet skin must have been crossed several times. Why this should happen is not very clear; usually it is supposed that the hailstones must have been lifted into the upper parts of the cloud several times in succession by a pulsatory up-current, but it is more probable that variations in the droplet- and raindrop-concentration within the cloud are responsible for this peculiar effect.

§ 7. THE DEVELOPMENT OF PRECIPITATION

(i) *Forms of Precipitation*

Considerable growth of a cloud particle must occur before it can gain a falling-speed which enables it to fall out of the parent cloud, survive evaporation in the unsaturated air beneath, and reach the ground as *precipitation*. Clearly the strength of the upward air-current producing the cloud to some extent determines the size of the precipitation elements. Widespread layer clouds are associated with the gentler upward air-movements of speed usually less than 0·5 m/sec., so that droplets with a radius approaching 80μ can fall out of them and approach the ground. Findeisen (1939 a) has estimated the distance a drop can fall through unsaturated air before completely evaporating, and has shown it to be proportional to almost the fourth power of its initial radius. In an atmosphere of 90% relative humidity

he finds this distance to be a few centimetres when the radius is 10μ, 150 m. when it is 100μ, and 42 km. when the radius is $1{,}000\mu$. Since (except in special circumstances) the bases of dense clouds lie at least a few hundred metres above the ground, a value for the radius of about 100μ may be regarded as a lower limit to the size of precipitation elements. Precipitation composed entirely of drops little larger than this characteristically falls in damp weather from layer clouds not far above the ground, and is called *drizzle*. Drizzle (and, in cold weather, small snowflakes or snow crystals) may fall from shallow clouds whose thickness amounts to only a few hundred metres. *Rain*, consisting of drops which are larger, having radii up to about 2 mm., but whose concentration is commonly much less, is produced by layer clouds some thousands of metres deep, such as are associated with depressions. The heaviest rains fall from convective (cumulus) clouds whose depth may reach 10 km. and which contain powerful up-currents of at least several metres per second. Precipitation from these clouds is sporadic, for their horizontal extent is only a few kilometres and the active life of individuals is less than an hour. These rains are therefore described as *showers*, and by his use of the word the meteorologist always implies that the parent clouds are of the convective type. Showers often contain hailstones (which in hot climates are sometimes as large as oranges), and indeed in cold weather the precipitation may reach the ground entirely as hail or snow.

(ii) *The Formation of Precipitation Elements*

Unless the concentration of cloud particles is extremely small it is impossible to produce precipitation elements by continued condensation from the vapour. The water available for the growth of cloud particles cannot exceed the 20–30 gm/kg. released by lifting damp tropical air to a great height; if this is divided equally amongst 10^9 droplets then each has a radius of less than 20μ, and to increase the radius to over 100μ the concentration must be reduced below about 7×10^6/kg. of air, that is, to less than $7/\text{cm}^3$. Such an extremely low concentration of droplets cannot be expected in view of the natural abundance of condensation nuclei.

On the other hand, ice crystals commonly occur without attaining even this concentration, and in the growth of snow from ice crystals continued condensation of vapour is of great importance. Ice clouds, even at levels where the low temperature restricts the total amount of water present to less than $1\ \text{gm/m}^3$, are often composed of particles large enough to have falling-speeds of nearly 1 m/sec. This is partly because the rare crystals appear only when the vapour pressure approximates to saturation over liquid water, so that beneath the condensation region there is a layer of air which, although not saturated with respect to liquid water, is nevertheless supersaturated with respect to ice. The crystals therefore continue to grow while they fall through this layer, which is often a kilometre deep and may be much more. In this way a few crystals draw on the vapour in a great volume of air and become much larger than the droplets of unfrozen clouds. Many ice clouds can thus be regarded as precipitation, with corresponding distinctive structures (Ludlam 1948).

(iii) *Growth of Coagulation Elements*

If these ice crystals fall into a supercooled cloud then collisions occur with the droplets, which become frozen. In a dense cloud containing many large droplets the collisions are so frequent that the crystals quickly become irregular, roughly spherical aggregates which develop into hailstones if the up-currents of air support

them in the cloud long enough. On the other hand, in cold weather even deep supercooled clouds often have very low liquid-water content, and although ice crystals falling through them occasionally encounter droplets, their growth proceeds predominantly by diffusion. Richly branched snow crystals are then developed in the manner described in § 6. Small fragments may break off these crystals and then themselves grow and produce further fragments, so that although initially the cloud may contain very few ice crystals, after a while the number may increase enormously, even to the point where coagulation between snow crystals begins and snowflakes are formed. Experience shows that large snowflakes, whose diameters may exceed a centimetre, are formed in cloud whose temperature lies close to 0° c., and it may be that the coagulation is greatly promoted when the surfaces of the snow crystals are wet, or when momentary melting and regelation occurs at the point of contact of colliding crystals. Even the largest snowflakes have fall-speeds hardly exceeding 1 m/sec., and so snow reaching the ground is characteristic of cold weather and falls from those layer clouds or decaying convective clouds which contain only small vertical air-currents.

Bergeron (1935) considered that most, if not all, raindrops are melted snowflakes. He examined a number of processes which might cause the coalescence of cloud droplets, including the effects of unlike electric charges on neighbouring drops, turbulence in the air, and the hydrodynamic forces between drops falling side by side. He concluded that all these were ineffective or much too slow, and that in the formation of rain the only factor of importance is the difference between the saturation vapour pressures over liquid water and over ice at temperatures below 0° c. Clouds which form in or which rise into regions where the temperature is below 0° c. contain a large number of supercooled droplets and a few ice crystals. In such *mixed clouds* the state of approximate water saturation represents a supersaturation with respect to ice of 10% at −10° c., and 21% at −20° c. The crystals therefore grow more rapidly than the droplets, which are subjected to supersaturations of the order 0·1%. Bergeron suggested that all raindrops originate as ice particles which ' grow copiously ' in this way and melt on approaching the ground, and therefore that all rainclouds must extend above the level of the 0° c. isotherm. He claimed to have conclusive proof in the observed behaviour of cumulus clouds, which do not produce showers until their summits reach above the 0° c. isotherm and ' glaciate ', becoming transformed into fibrous ice clouds. This transformation is illustrated in Plate 13. Bergeron's hypothesis was vigorously supported and developed by Findeisen, who on the basis of aerological observations over Germany stated categorically (1939 a) that all rain composed of large drops originates in falls of snow or hail. Findeisen emphasized that once an ice crystal in a dense supercooled cloud has grown appreciably larger than the droplets, its growth is greatly accelerated by coagulation with them following collisions due to its greater falling-speed.

A similar colliding process might be expected to lead to the continued growth of the larger of the droplets in wholly liquid clouds, which have never been found quite uniform, but has been regarded as a slow and inefficient mechanism of precipitation-release capable of causing only falls of drizzle (Bergeron 1949). Nevertheless when Findeisen (1939 b) calculated the radii of the drops which would thus be expected to fall from a typical layer cloud he found values of 0·4 and 0·7 mm. for cloud thicknesses of 1,000 and 2,000 m. respectively. He rejected this result as contrary to experience, and believed his theory to be in error

in that it assumed that a falling drop collects all the droplets in its path, none being brushed aside. Reviewing the subject in 1941 Simpson emphasized that the belief that raindrops necessarily originate as ice particles rests ultimately on statements that rain is never observed to fall from pure droplet-clouds, even when they are several thousand metres thick. Simpson would not accept these statements, and suggested that copious rain could be produced by wholly liquid clouds providing only that the original droplets were of different sizes so that the coagulation process could start, and that the cloud was of sufficient depth.

Recently careful observations have accumulated which show beyond all doubt that in tropical regions heavy rain falls from clouds which are entirely beneath the level of the 0° c. isotherm (Kotsch 1947, Hunt 1949, Schaefer 1949 b, Virgo 1950, Smith 1951). In these regions this level may be above 5 km., whereas in our latitudes it is often below 2 km., and here equally deep clouds commonly reach heights where the temperature is below −20° c., so that their tops would certainly contain ice crystals. It does not follow, however, that their presence is essential to the production of rain. The present evidence does suggest that the rain which falls from thick layer clouds in extra-tropical regions is probably entirely initiated by the appearance of ice crystals in their upper, supercooled parts;* on the other hand it seems equally probable that some rain would fall from such thick clouds even in the absence of the ice crystals, and that more critical observation may occasionally show this to occur (see, for example, Kinshott 1950).

(iv) *The Multiplication of Raindrops by a Chain Reaction*

Another problem is presented by the very rapid production of heavy showers from convective clouds. Rain is often seen to fall from the base of such clouds within twenty minutes of their towering into prominence. It is of great importance that within many of these clouds there prevail up-currents of speed comparable with or greater than the maximum terminal velocity of a raindrop. Falling drops with radii greater than 1 mm. suffer a distortion which becomes ever more pronounced until at an equivalent radius of about 3 mm. and a falling speed of about 10 m/sec. the distortion becomes so great that any further increase in the mass results in the disruption of the drops. Experiment (Lenard 1904, Blanchard 1949) shows that the exact stage at which this breakage occurs and the nature of the fragments vary considerably, especially with the degree of micro-turbulence in the air stream, but that usually several droplets of about 1 mm. radius and relatively large numbers of much smaller droplets are produced. This implies that raindrops growing by coagulation in a cloud containing an up-current of about 10 m/sec. cannot descend, and that their numbers are constantly increasing. Eventually however, their accumulation must overcome the buoyancy of the air which sustains the up-current, so that they are released as a heavy shower which by its drag and by chilling due to its partial evaporation may

* It appears that ice nuclei occur in sufficient abundance to account for the rates at which raindrops reach the ground during steady rain (Palmer 1949). Here it must be remarked that the number of effective ice nuclei may be greatly increased by the separation of fragments from snow crystals. Findeisen (1940, 1943) found that air flowing over a growing or evaporating ice surface carries away with it small 'splinters' of ice, whose radius is about 20μ. This splinter formation does not occur if the ice surface is glassy or part of a simple crystal, but occurs over ice particles which are aggregates of small crystals. The separation of the ice splinters appears to be associated with the collapse of outgrowing dendrites, but the physics of the process is very imperfectly understood. Findeisen recognized its importance as a source of ice nuclei inside clouds, and further believed it to be responsible for the generation of thunderstorm electricity, his experiments showing that the production of splinters is associated with the separation of appreciable electric charges.

generate an accompanying down-current as strong as the original up-draught. Recently Langmuir (1948) has examined the critical conditions under which the fragments of a drop disrupting in up-currents of less than 10 m/sec. themselves reach break-up size before they have fallen back again to their level of origin, and therefore under which a *chain reaction of raindrop multiplication* can occur. He advances the coagulation theory by calculating the fraction E ('collection efficiency') of the droplets in the path (i.e. the projected cross section), of a spherical collector with which collision (and presumably coalescence) occurs, in terms of the respective radii and relative falling-speed. His formulae show that for a given size of the smaller droplet no collision occurs until the radius of the collector reaches a critical value, and that beyond this point the efficiency of catch E of the collector increases with its radius, becoming about 0·9 for a large raindrop falling through a cloud composed of droplets of normal size. The formulae are based on the assumption that the collecting drop is large compared with the cloud droplets, and the all-important problem of the initiation of the coagulation process amongst nearly uniform droplets remains unsolved, but at least valuable conclusions can be drawn on the way in which the precipitation process develops once it has been started.

It straightway appears that the chain reaction is a very efficient mechanism of precipitation release. The growth of a drop by coagulation may be described by the relations

$$\frac{dm}{dt} = 4\pi r^2 \rho \frac{dr}{dt} = E\pi r^2 v W \qquad \text{.....(13)}$$

or

$$\frac{dr}{dt} = \frac{EvW}{4\rho}, \qquad \text{.....(14)}$$

where dm/dt is the rate of increase of mass, W is the concentration of water in the form of the smaller droplets, v is the relative falling-speed of the collecting drop, and ρ is the density of the liquid. In the central parts of a large convective cloud of temperate latitudes W reaches about 4×10^{-6} gm/cm^3; for drops of radius 1 to 3 mm. E is about 0·9 and v (at these levels) averages about 10^3 cm/sec., so that over this interval of radius

$$\frac{dr}{dt} = \frac{0{\cdot}9 \times 10^3 \times 4 \times 10^{-6}}{4} = 9 \times 10^{-4} \text{ cm/sec.}$$

Thus under these conditions the larger fragments of a breaking raindrop grow to disruption size in less than four minutes, and since a 3 mm. drop may break into some ten fragments the concentration of raindrops increases by a factor of about ten every four minutes. If the cloud contains an up-draught of about 10 m/sec. and drops of 1 mm. radius are introduced into its upper parts in a concentration of 1/m^3, then after fifteen minutes these would have been replaced by large raindrops numbering nearly 1,000/m^3 and representing a liquid water concentration of over 20 gm/m^3, which would certainly be sufficient to destroy the up-draught and release a heavy shower.

The question now arises whether there is any natural mechanism of introducing a few large drops into the upper parts of such clouds; near the cloud base the up-currents and concentrations of cloud water are so much reduced that a chain reaction cannot be maintained, the larger drops falling out of the cloud before reaching break-up size. The observation that cumulus tops ordinarily glaciate before showers occur suggests that large ice particles grown in the manner

postulated by Bergeron might produce the required drops by melting when they fall below the level of the 0° c. isotherm. On the other hand, in the tropics heavy showers are produced by cumulus clouds which fail to reach this level, and in these some other mechanism must be active.

The nuclei containing 10^{-10} gm., or even more, of salt, which have been found at all heights to above 1,000 m. over the sea near Bermuda (Woodcock and Gifford 1949) and Florida (Woodcock 1950), in concentrations of the order of 1/litre, have at a relative humidity of 100% radii of 80 μ or more, and are able to introduce drops having radii of 30 μ or rather more into the clouds of these regions.

Calculations by Ludlam (to be published later) show that in the dense cumulus of tropical climates droplets of radius about 30 μ grow sufficiently rapidly by coagulation to reach the break-up size before they have been lifted above the 0° c. level.* Large salt nuclei could thus bring about shower production from clouds not reaching this level. No investigations have yet been made to determine whether they exist well inland as well as over the sea; it is probably significant that the observations of rain from non-freezing clouds have been made predominantly over the oceans or near coasts. Measurements of the concentration of the large nuclei will not be easy, since only 1/m³ or even less may efficiently initiate a chain reaction of raindrop multiplication.

Away from tropical regions the 0° c. level falls so much lower and the liquid water content of clouds below it becomes so reduced that it is not possible to produce rain in this way from cumulus whose tops are not supercooled However, stratiform clouds which exist a relatively long time are doubtless aided by the large nuclei in the production of drizzle, and their growth into raindrops must tend to occur within cumulus in whose tops ice crystals appear and which are therefore provided with an alternative mechanism of rain production.

It is now important to consider whether the appearance of ice crystals is *essential* to the formation of rain in those convective clouds of temperate climates which are deep enough to support a chain reaction, and which therefore rise above the 0° c. level. With this in view Ludlam (to be published later) has estimated the course of growth of both ice crystals and the large condensation nuclei in these clouds. Figure 9, for example, describes the growth in a vigorous cloud having a base temperature of 10° c. and containing an up-current increasing from 5 m/sec. at the base to 10 m/sec. at 3·5 km. and above. In this up-current the first ice nuclei are presumed to become active at a temperature of −5° c.; after about five minutes the ice crystals have grown into hailstones which begin to descend through the up-current. Nearly 1,000 sec. after the air containing the ice nuclei rose through the cloud base the hailstones have diameters of over a centimetre, and their surfaces now become wet at heights where the outside temperature is below −20° c. Beyond this stage the hailstones acquire so much supercooled water that the rate at which latent heat of fusion can be disposed of to the environment limits the rate of freezing, and liquid water accumulates on their surfaces (Ludlam 1950). After a short while some of the excess water may be shed, presumably as droplets with radii of the order of a millimetre. In this way the first formed hailstones are able to introduce into the cloud large drops, which commence a chain reaction, even before they have fallen below the 0° c. level and have begun to melt. However, we may now enquire into the

* *Note added in proof.* Since this report was written a description of similar calculations has been published by E. G. Bowen (1950). In this important paper Bowen also describes radar evidence of the development of rain within cumulus clouds *before* they have risen to the 0° c. level.

history of large condensation nuclei presumed to be present in the same air which held the efficient ice nuclei. These are considered to have produced droplets of radii 30, 40 and 50 μ when the air has reached 2 km., and the solid lines in the Figure show their subsequent growth up to the stage at which their radii become 1 mm. Unfortunately, the earlier stages when the radii are less than about 100 μ are inexactly followed, because the growth is assumed to occur with an 'efficiency of catch' calculated from Langmuir's formulae. Nevertheless the error is not likely to be serious in the case of the 50 μ droplet, and probably results in a calculated growth which is too slow rather than too fast. With this in mind we see that the large nuclei produce 1 mm. droplets at about 6 km., and nearly 300 sec. *before* the arrival of the first formed hailstones from above. The large nuclei are therefore capable of forestalling the ice particles in initiating a chain reaction of raindrop multiplication, and indeed if they do so they must actually

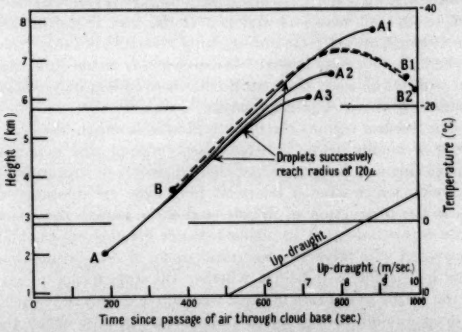


Figure 9. Growth of ice particles and droplets in a cumulus cloud with base at 890 m. and 10° c., and containing an up current increasing from 5 m/sec. at the cloud base to 10 m/sec. at 3,500 m. At A droplets of radius 30, 40 and 50μ are introduced into rising cloud-air, and these droplets follow the solid lines, reaching a radius of 1 mm. at the points A1, A2 and A3 respectively. Ice nuclei in the same air are assumed to act when the temperature has fallen to $-5°$ c., at the point B; the ice particle following the upper broken line is growing with a density of 0·3 gm./cm^3, and one following the lower broken line has a density 0·6 gm/cm^3. At the points B1, B2 the ice particles which have now become hailstones of radii 1·0 and 0·6 cm. begin to acquire liquid skins and may therefore soon shed large drops into their wakes.

hinder the growth of many crystals by removing liquid water from the air rising up to feed them. In a cloud having the structure described there are no particles capable of becoming precipitation elements until the cloud tops have reached about 4·5 km., where the temperature is $-10°$ c.; if at this point the cloud growth were interrupted large droplets are present which could survive the evaporation of the summits (which is usually observed within a few minutes) and infect newly rising cloud masses, so that the development of a shower is already possible. If the cloud growth is halted at successively later stages, shower production (and the chain reaction of raindrop multiplication) becomes increasingly probable, and it is almost certain that traces of ice crystals (glaciation) will be observed at the cloud summits. More detailed evidence than visual observation, however, is needed to establish whether the appearance of ice crystals is more important than the presence of large condensation nuclei. The indirect evidence at present available suggests that even in temperate latitudes the latter may anticipate the

crystals in initiating the shower-rains. Amongst this evidence is the knowledge that frozen precipitation (hail) does not occur in all showers, even when in rather cold weather it might be expected to reach the ground before melting, and the continuous manner in which the critical summit-temperature which must be reached before cumulus become cumulo-nimbus (shower clouds) decreases with lowering base-temperature.*

(v) *The Artificial Stimulation of Precipitation*

Knowledge of the exact manner in which natural showers develop is needed before the possibilities of producing artificial rain can be assessed. It has been supposed that the introduction of artificial ice nuclei into the supercooled tops of cumulus which fail to produce precipitation would remedy a natural absence of efficient nuclei and cause shower-development. The artificial nuclei are readily introduced by dropping into the cloud pellets of solid carbon dioxide, or from the ground by disseminating large numbers of silver iodide particles, some of which are drawn into convective clouds by the up-currents producing them. Experiments of this nature have been tried in several countries recently since the techniques were developed by Langmuir and his associates in America. Findeisen had begun similar work in Europe already in 1944. The majority of these attempts to induce rain have apparently accelerated the dissipation of the cloud rather than resulted in any appreciable precipitation (Coons *et al.* 1949, Smith 1949), but occasionally it has been reported that the 'seeded' clouds have swelled much above their neighbours with the production of heavy rains (Kraus and Squires 1947), see Plate 14. It seems that carbon dioxide pellets (which acquire a coat of ice while falling through cloud) may promote chain reactions of raindrop multiplication by introducing large droplets below the 0° c. level rather than by producing crystals above it, an explanation advanced by Langmuir for the successful seeding in this way of a Hawaiian cloud wholly below the 0° c. level. This method of seeding might therefore fail consistently in cold or weak clouds incapable of maintaining a chain reaction. It has been recognized that in order to produce precipitation by ice-nucleus seeding the concentration of nuclei must approximate to some critical value of the order of $1/\text{cm}^3$, numbers much in excess of this resulting in the production of a dense ice cloud with small particles, and numbers much less not producing sufficient precipitation elements (unless a chain reaction is started). The great difficulty of controlling the concentration has recently led to the practice of producing very large numbers of nuclei—'overseeding'—in the hope of causing abnormal developments easily recognized and perhaps of value in throwing light on natural precipitation development. If the whole of a supercooled cloud in the region of the -10°c. level is in this way transformed into a pure ice cloud, the associated release of latent heat may be sufficient to warm the cloud air through 4° c. or more, thereby accelerating its upward motion and prolonging its rise. Such a transformation may account for the occasional towering of seeded clouds, the subsequent release of heavy rain perhaps being accomplished naturally in the air following the rise of infected region. The seeding of clouds with water droplets has also been attempted, so far with inconclusive results: Schaefer (1949 b), having prepared thus to seed cumulus near Puerto Rico, found that these clouds produced rain naturally, although their tops lay much below the 0° c. level.

* Ludlam, unpublished. This relation is also indicated in data described by Schwerdtfeger (1948).

A great trouble in these experiments with convective clouds is the difficulty of identifying the seeded cloud over a period of up to half an hour, and of judging whether or not the observed developments would have occurred naturally. (Experience has shown the use of centimetric radar to be an invaluable aid in the examination of precipitating clouds, allowing the determination of the location and vertical extent of regions containing droplets or crystals or radius larger than about 0·5 mm. within 50 to 100 miles of the apparatus. The possibilities of combining this technique with the visual examination of individual clouds have not yet been fully exploited.)

Although we have some knowledge of the microphysical aspects of the development of precipitation it must be emphasized that the macrophysical structure of clouds is equally important and perhaps less well understood. It will be appreciated that there is much interesting research to be done in both directions, and that considerable progress will be needed before a complete understanding of the production of drizzle, rain, snow and hail is gained, and before any chance of controlling their occurrence and intensity can be realized.

§8. CONCLUSION

In every droplet-cloud there are coagulation processes at work tending to produce larger and larger drops. Within uniform clouds the coagulation may be brought about by Brownian motion or the hydrodynamic attraction between droplets falling side by side; subsequently it is accelerated by coalescence resulting from collision of droplets of different sizes due to their relative fall-speeds. Clouds forming at temperatures below 0° C. and indeed down to at least −40° C. are usually at first composed predominantly of supercooled droplets, and in them similar effects occur, although it may then be almost exclusively the relatively rare ice particles whose continued growth is favoured. Thus in all natural clouds of appreciable water-content there is an ever present tendency for the production of precipitation; at first, however, the responsible processes operate extremely slowly, particularly in clouds of uniform droplets, so that the development of precipitation requires the maintenance of the cloud for some considerable period.

For this reason the large-scale physics of the cloud acquires a crucial importance. The life of individual cloud particles may be limited to less than a minute, as in the stationary wave-clouds formed in the lee of hills, or it may be allowed to extend over a day or longer, as in the interior of thick layer clouds often found over the oceans. It has also to be recognized that the life of a cloud as a whole is sometimes longer than that of its individual parts, which here is the important factor. Thus isolated cumulus clouds, although sometimes identifiable throughout an hour or more, are assemblages of smaller cloud masses which are continually dissipated near the sides and summit, and replaced near the base, and whose life extends over perhaps only a few minutes. However, if large droplets (of radius about 100 μ) are formed, then during such a period they can fall several hundred metres, thus leaving the air in which they first grew and entering fresh cloud masses; their further growth may then be controlled by the strength of the upward air current which delays their fall through the central parts of the cloud bulk.

The meteorologist, having determined these large scale features of particular clouds, has next to consider how rapidly the growth of the cloud particles into precipitation elements may proceed, and whether this growth can be prolonged to the stage in which precipitation reaches the ground. At present it is not

possible to make accurate calculations owing to the absence of reliable data on several of the processes involved in droplet growth.

The rapidity with which the coagulation of droplets proceeds depends markedly on their initial size and inhomogeneity. The initial size distribution is largely determined by the characteristics of the available condensation nuclei. It is probable that particles of a great variety of substances serve as nuclei in the atmosphere, but more detailed information is required concerning their concentration, efficiency and size, and on the way in which these properties vary from region to region both near the ground and in the upper air. In particular it appears that sea-spray, while supplying the bulk of the nuclei used in clouds formed over and near the oceans, is of even greater importance as a source of relatively big but rare hygroscopic nuclei which produce very large droplets in the clouds and thus efficiently initiate the coagulation process. The difficulties attending the sampling of atmospheric nuclei are emphasized by the calculation that such large nuclei may be of critical importance even when present in concentrations of less than $1/m^3$. The accurate determination of the chemical and physical nature of bodies as small and numerous as the nuclei present in such volumes of air seems a hopeless task. The most profitable approach to the establishment of their behaviour must lie in the examination of drop-size spectra in clouds produced by the expansion (at atmospheric rates) of representative samples of air, supplemented by techniques such as that used by Woodcock and Gifford for determining the concentration of large and rare nuclei. More careful and thorough sampling of the droplets found in natural clouds is also desirable; relatively rare large droplets must be sought for, and the observations must be related to the large-scale features of the cloud structure. Here there is an urgent need for a new method allowing the rapid analysis of the size distribution in droplet samples, to replace the present tedious technique of visual counting. The development of an instrument capable of making reliable measurements (with errors less than 10%) of the total liquid water content within clouds would also be of immediate value.

A further obstacle in the development of the theory of rain production is our ignorance of the laws governing the coalescence of droplets of nearly equal size. It is not even known whether coalescence always follows contact. It may be possible to obtain some information experimentally, but it will probably be very difficult to develop a comprehensive theory.

In those supercooled clouds within which the ice phase appears an analogous set of problems arises. Again there is a need for the basic data on the nature, efficiency and concentration of the active nuclei, and here the experimental difficulties are aggravated by the rarity of ice nuclei and the troubles encountered in working at sub-freezing temperatures with restricted volumes of air. It seems that the efficiency of ice nuclei may in some way be affected by their age and history, and by the rate of expansion of the air in which they float. Here also the microphysics of their behaviour may well be so complex that a purely empirical investigation of their occurrence and properties is most likely to be immediately profitable. The concentrations in which ice nuclei represent the most efficient agents of precipitation release, and the extent to which natural deficiencies in their number are remedied by the 'splintering' of ice particles, are matters on which only vague estimates have so far been made, but which must be of the greatest importance in the production of snow and rain from stratiform clouds.

In cumuliform clouds capable of supporting chain reactions of raindrop multiplication the precise initial concentration of ice nuclei and the splintering process may not be of critical concern, but the development of the ice phase within these clouds must modify their structure and behaviour, and may also be important in the generation of thunderstorm electricity.

The manner in which large droplets, or ice particles, grow into precipitation elements must be accurately understood before the possibilities of the artificial control of precipitation can be properly appreciated. Belief in the exclusive importance of ice particles in shower formation, and the discovery of methods of producing cheaply vast numbers of ice nuclei more efficient than those occurring naturally, have led to widespread attempts to provoke or increase rainfall by the dissemination of these artificial nuclei. Some of these attempts have been successful, but the very great difficulty of correctly analysing their results and the publication of exaggerated claims have caused many meteorologists to regard these projects with scepticism or even amusement. This is unfortunate; the release of precipitation, the dissipation of cloud and the occurrence of thunderstorms are all phenomena intimately related to the microphysics of the clouds, and it has long been recognized that by some appropriate stimulation requiring the expenditure of very little energy the microphysical processes might be sufficiently helped or retarded to cause significant changes in weather. The economic advantages to be gained by even partial control of rainfall and sunshine are so enormous that great effort may justifiably be spent in attempting their realization. It seems certain, however, that before success can be achieved much more study will be needed of those aspects of cloud physics which have been discussed in this article, and perhaps also of others which have not yet been appreciated.

REFERENCES

AANENSEN, C. J. M., 1948, *Met. Mag.*, **77**, 17.

BARENDRECHT, H. P., 1896, *Z. phys. Chem.*, **20**, 234.

BECKER, R., and DÖRING, W., 1935, *Ann. Phys., Lpz.*, **24**, 719.

BERGERON, T., 1935, *Proc.-Verb. de l'U.G.G.I., Lisbon*, 1933 ; 1949, *Tellus*, **1**, 32.

BLANCHARD, D. C., 1949, *G.E.C. Occasional Report, No.* 7, Project Cirrus.

BOWEN, E. G., 1950, *Aust. J. Sci. Res.* A, **3**, 193.

BOYLAN, R. K., 1926, *Proc. R. Irish Acad.*, A, **37**, 58.

BRICARD, J., 1939, *La Météorologie*, **15**, 83.

BURTON, W. K., and CABRERA, N., 1949, *Crystal Growth* (London: Faraday Society), p. 40.

BURTON, W. K., CABRERA, N., and FRANK, F. C., 1951, *Proc. Roy. Soc.* A, in the press.

BYERS, H. R., and BRAHAM, R. R., 1948, *J. Met.*, **5**, 71.

COONS, R. D., JONES, E. L., and ROSS GUNN, 1949, *U.S. Weather Bur., Res. Papers*, **31, 33.**

COSTE, J. H., and WRIGHT, H. L., 1935, *Phil. Mag.*, Ser. 7, **20**, 209.

CWILONG, B. M., 1947 a, *J. Glaciology*, **2**, 53; 1947 b, *Proc. Roy. Soc.* A, **190**, 137; 1947 c, *Nature, Lond.*, **160**, 198; 1949, *Ibid.*, **163**, 727.

DESSENS, H., 1946, *Ann. Géophys.*, **2**, 276, 343 ; 1947, *La Météorologie*, p. 321; 1949, *Quart. J.R. Met. Soc.*, **75**, 23

DIEM, M., 1942, *Ann. Hydrogr., Berlin*, **70**, 142.

DOLOZEL, E. J., CUNNINGHAM, R. M., and KATZ, R. E., 1946, *Bull. Amer. Met. Soc.*, **27**, 261.

DORSEY, N. E., 1938, *Bur. Stand. J. Res., Wash.*, **20**, 799 ; 1948, *Trans. Amer. Phil. Soc.*, **38**, 248.

FINDEISEN, W., 1937, *Met. Z.*, **54**, 377; 1938, *Ibid.*, **55**, 121; 1939 a, *Ibid.*, **56**, 365; 1939 b, *Ibid.*, **56**, 453 ; 1940, *Ibid.*, **57**, 201 ; 1942, *Ibid.*, **59**, 349 ; 1943, *Ibid.*, **60**, 145.

FINDEISEN, W., and SCHULZ, G., 1944, *Forsch. u. Erfahrb. Reichs. Wetterdienst, Berlin*, A, No. 27.

FOURNIER D'ALBE, E. M., 1949, *Quart. J.R. Met. Soc.*, **75**, 1.

FRANK, F. C., 1949, *Crystal Growth* (London : Faraday Society), p. 48.

FRÖSSLING, N., 1938, *Gerlands Beitr., Lpz.*, **52**, 170.
GLAUERT, M., 1940, *Rep. Memor. Aero. Res. Comm., Lond.*, No. 2024.
HOLLSTEIN, FRL., 1947, *Die Eisphase in der Atmosphäre, Rep. and Trans.*, No. 716 (Volkenrode), Ministry of Supply (A).
HOSLER, C. L., 1950, *Trans. Amer. Geophys. Un.*, **31**, 707.
HOUGHTON, H. G., and RADFORD, W. H., 1938, *Pap. Phys. Oceanogr. Met.*, **6**, No. 4.
HOWELL, W. E., 1949, *J. Met.*, **6**, 134.
HUMPHREYS, W. J., 1929, *Physics of the Air* (London and New York; McGraw-Hill), p. 516.
HUNT, T. L., 1949, *Met. Mag.*, **78**, 26.
JUNGE, C., 1935, *Gerlands Beitr., Lpz.*, **46**, 108 ; 1936, *Met. Z.*, **53**, 186.
KINSHOTT, E. J., 1950, *Met. Mag.*, **79**, 205.
KÖHLER, H., 1922, *Geofys. Publ.*, **2**, No. 6; 1941, *Nova Acta Reg. Sci., Uppsala*, **12**, No. 6.
KOTSCH, W. J., 1947, *Bull. Amer. Met. Soc.*, **28**, 87.
KRASTANOW, L., 1940, *Met. Z.*, **57**, 357 ; 1941, *Ibid.*, **58**, 37 ; 1943, *Ibid.*, **60**, 15; 1944, *Ibid.*, **61**, 19.
KRAUS, E. B., and SQUIRES, P., 1947, *Nature, Lond.*, **159**, 489.
LAFARGUE, C., 1950, *Quart. J. R. Met. Soc.*, **76** (Centenarv Proceedings).
LANGMUIR, I., 1944, *G.E.C. Report*, 'Supercooled Water Droplets in Rising Currents of Cold Saturated Air', Parts I and II ; 1948, *J. Met.*, **5**, 175.
LENARD, P., 1904, *Met. Z.*, **21**, 249.
LUDLAM, F. H., 1948, *Quart. J. R. Met. Soc.*, **74**, 39; 1950, *Ibid.*, **76**, 52.
MASON, B. J., 1950 a, *Quart. J. R. Met. Soc.*, **76**, 59; 1950 b, *Ibid.*, **76** (Centenary Proceedings); 1951, *Ibid.*, **77**, in the press.
MAY, K. R., 1950, *J. Sci. Instr.*, **27**, 128.
MEYER, J., and PFAFF, W., 1935, *Z. anorg. Chemie*, **224**, 305.
PALMER, H. P., 1949, *Quart. J. R. Met. Soc.*, **75**, 15.
PAUTHENIER, M., and BRUN, E., 1940, *C. R. Acad. Sci., Paris*, **211**, 295.
RAU, W., 1944, *Schr. deutsch. Akad. Luft., Berlin*, **8**, 65.
REGENER, E., 1941, *Schr. deutsch. Akad. Luft., Berlin*, H **37**, 15.
ROSSMANN, F., 1948, *Met. Rdsch.*, **1**, 193.
SANDER, A., and DAMKÖHLER, G., 1943, *Naturwissenschaften*, **31**, 460.
SCHAEFER, V. J., 1946, *Science*, **104**, 457 ; 1948, *Bull. Amer. Met. Soc.*, **29**, 175; 1949 a, *J. Met.*, **6**, 283 ; 1949 b, *G.E.C. Occas. Report*, No. 12, Project Cirrus; 1950, *Ibid.*, No. 20, Project Cirrus.
SCHULZ, G., 1948, *Met. Rdsch.*, **1**, 237.
SCHUMANN, T. E. W., 1938, *Quart. J. R. Met. Soc.*, **64**, 3; 1940, *Ibid.*, **66**, 195.
SCHWERDTFEGER, W., 1948, *Met. Rdsch.*, **1**, 453.
SHEPPARD, P. A., 1947, *Sci. Progr.*, **35**, 490.
SHIRATORI, K., 1934, *Mem. Fac. Sci. Taihoku Univ.*, **10**, 175.
SIMPSON, G. C., 1941 a, *Quart. J. R. Met. Soc.*, **67**, 99 ; 1941 b, *Ibid.*, **67**, 163.
SMITH, E. J., 1949, *Aust. J. Sci. Res.* A, **2**, 78; 1951, *Quart. J. R. Met. Soc.*, **77**.
SMITH-JOHANNSEN, R., 1948, *Science*, **108**, 652.
STÖRMER, C., 1948, *Weather*, **3**, 13.
VIRGO, S. E., 1950, *Met. Mag.*, **79**, 237.
VOLMER, M., 1939, *Kinetik der Phasenbildung* (Dresden and Leipzig: Steinkopff).
VOLMER, M., and ESTERMANN, I., 1921, *Z. Phys.*, **7**, 13.
VOLMER, M., and FLOOD, H., 1934, *Z. phys. Chem.* A, **170**, 273.
VONNEGUT, B., 1949 a, *Rev. Sci. Instrum.*, **20**, 110; 1949 b, *Chem. Rev.*, **44**, 277.
WALL, E., 1947, *Wiss. Arb. der Franz. Zone*, **1**.
WEGENER, A., 1911, *Thermodynamik der Atmosphäre* (Leipzig: J. A. Barth); 1920, *Met. Z.*, **37**, 8.
WEICKMANN, H., 1942, *Forsch. Bericht.* 1730 *Deutsch. Luft., Berlin* ; 1947, *Die Eisphase in der Atmosphäre, Rep. and Trans.*, No. 716 (Volkenrode), Ministry of Supply (A).
WIGAND, A., 1913, *Met. Z.*, **30**, 10.
WOODCOCK, A. H., 1950, *J. Met.*, **7**, 161.
WOODCOCK, A. H., and GIFFORD, M., 1949, *J. Mar. Res.*, **8**, 177.
WRIGHT, H. L., 1936, *Proc. Phys. Soc.*, **48**, 675; 1939, *Quart. J. R. Met. Soc.*, **65**, 411; 1940, *Ibid.*, **66**, 66.

ANGULAR CORRELATIONS IN NUCLEAR REACTIONS

By MARTIN DEUTSCH

Laboratory for Nuclear Science and Engineering,
Massachusetts Institute of Technology, Cambridge, Mass.

CONTENTS

	PAGE
§ 1. Introduction	196
§ 2. The basic physical processes	197
(i) Experiments involving conservation of momentum	197
(ii) Experiments testing laws of interaction	197
(iii) Experiments to study properties of nuclear states	198
(iv) Elementary theory of angular correlations	199
§ 3. Experimental techniques	204
(i) Angular distributions	204
(ii) Angular correlations	210
(iii) Experiments involving more than one angle	213
§ 4. Experimental results	214
(i) Induced reactions	214
(ii) Radioactive decay	220
References	224

ABSTRACT. Experiments involving the measurement of angular distributions and correlations of nuclear radiations fall into three groups: those involving simply conservation laws, studies of nuclear interactions and investigations of the properties of nuclear states. Elementary theoretical considerations and typical or important experimental results are presented with particular emphasis on the third group. When all the states involved are 'pure' states with defined angular momentum and parity the correlations are always polynomials in *even* powers of cos θ only whose coefficients depend on the angular momenta of the states and the nature of the radiations. This condition of pure states is satisfied for all radioactive processes and generally for nuclear reactions at a resonance. Consequently it has been possible to deduce the angular momentum quantum number of a number of excited states from observed angular correlations between gamma-rays and alpha-rays, protons, and gamma-rays respectively. In non-resonant induced reactions a compound nucleus is frequently formed with so high an excitation that overlapping states of opposite parities contribute to the process. In this case odd powers of the cosine may appear and the angular distributions are not necessarily symmetrical about 90°. This is particularly true in reactions of the type (d, p) or (d, n) and, probably, (d, t) in which no true compound nucleus is formed at all. The distribution in these reactions is usually strongly forward.

§ 1. INTRODUCTION

THERE is hardly a problem in experimental nuclear physics whose solution does not involve the measurement of an angle. Even the discussion of all experiments concerned only with dependence of the intensity of some nuclear radiation on an angle would approach a textbook of nuclear physics.

In this report we shall deal mostly with experiments designed to study properties of nuclear energy states. Even with this restriction it is not possible to present all important techniques and results. Some of the more interesting examples are quoted but the selection is necessarily somewhat arbitrary. Emphasis

has been placed on giving a broad view of the physical problems and experimental methods involved rather than on distributing credit in proportion to original achievement.

§2. THE BASIC PHYSICAL PROCESSES

From the point of view of the fundamental physical processes involved, most experiments on angular distributions and correlations fall in one of three categories: (i) correlations based on the laws of conservation of linear momentum; (ii) experiments designed to test the law of interaction between nuclei and the emitted or absorbed rays; (iii) experiments designed to study the properties of the energy levels of the nuclei involved. The dividing line between the second and third of these is not very sharp. In many experiments, such as the study of angular correlations between beta- and gamma-rays, both problems are investigated. In other cases broad or overlapping virtual levels of the nuclei appear and it becomes a matter of convenience whether one wishes to describe the process in terms of a scattering problem or of well-defined intermediate states. The main subject of this report is covered by the third category of experiments, but some results coming properly in the second group are also discussed.

(i) *Experiments involving Conservation of Momentum*

A classic example of this type of experiment is the investigation by Compton and Simon (1925) of the angular correlation in x-ray scattering between incoming quantum, scattered quantum and recoil electron which strongly confirmed the corpuscular character of the process by testing the validity of the conservation laws. Recently the angular correlation between the two quanta emitted upon annihilation of a positron and an electron has been investigated by Beringer and Montgomery (1942) and by deBenedetti, Cowan and Konneker (1949) to determine the distribution in positron velocity at the time of annihilation. If the positron has appreciable velocity, conservation of momentum requires that the two quanta make an angle slightly different from 180° with each other.

(ii) *Experiments testing Laws of Interaction*

A classic example is the establishment of the scattering law for alpha-particles which led to the discovery of the nuclear atom. In recent years extensive experiments have been carried out to investigate the nature of forces between nucleons by studies of neutron–proton and proton–proton scattering. These experiments, and related ones involving other very light nuclei, are of great fundamental importance. They form an extensive subject by themselves and will not be discussed in this report. Other examples, also beyond the scope of this report, might be the test of the Klein–Nishina formula and the investigations of electron–electron scattering.

Very closely related to scattering experiments of the kind just discussed is the case of the *simultaneous* emission of two particles as distinguished from the case of the *successive* emission involving a well-defined intermediate state. In fact, in a formal way, the theory of such simultaneous emission can always be transformed into a scattering problem. Examples of this are the angular correlation between the components of internally converted pairs (Devons and Lindsey 1949) and the correlation between electron and neutrino—or more precisely between electron and recoil nucleus—in beta decay (Sherwin 1948, 1949, Allen, Paneth and Morrish 1949).

(iii) *Experiments to study Properties of Nuclear States*

In these experiments the dependence on angle of the successive emission of two rays or particles (or the absorption of one and emission of another) is investigated and interpreted on the assumption that between the processes the nucleus is in a well-defined intermediate state to which an angular momentum quantum number and parity, and possibly other quantum numbers, can be assigned.

During the last two decades methods of establishing the excitation energies of nuclear states have made tremendous strides. In the case of radioactive processes techniques for determining the energies of alpha-, beta- and gamma-rays have been refined, mainly by magnetic deflection methods but also by special procedures, to attain great precision, although the accuracy of optical spectroscopy cannot yet be matched. Ingenious methods have been developed to construct the energy level schemes once the radiations are known (Maier-Leibnitz 1942, Mitchell 1950). Most of these methods are based on coincidence techniques used to establish which radiations are emitted in cascade in a given decay. Other excited states are found from the energies of particles given off by unstable nuclei produced in the bombardment with swift charged particles or neutrons or, to a lesser extent, slow neutrons or electromagnetic radiation (Hornyak 1950, Hafner 1950).

With an extensive catalogue of excitation energies thus developing rapidly, interest has turned towards investigating other properties of excited states, particularly their angular momenta and parities and, if possible, their electric and magnetic moments. This desire has been enhanced on the one hand by the great progress in the study of these properties for ground states of nuclei through atomic and molecular beams (Kellogg and Millman 1946) and radio-frequency spectroscopy (Gordy 1948), and on the other hand by the recently revived interest in simple models for the structure of heavier nuclei, which seems to offer some hope of interpreting the results of the experiments (see, for example, Mayer 1950 for literature).

Considerable information can be gained from the relative intensities of several competing modes of transition between the same states. In the case of electromagnetic transitions the coefficients of internal conversion in the K, L and M shells or by pair creation can be compared with the theoretical predictions for multipole transitions of various types (Hebb and Nelson 1940, Dancoff and Morrison 1939, Drell 1949, Rose 1949, Rose, Goertzel, Spinrad, Harr and Strong 1949). These are in turn related to the properties of the nuclear states by well-known selection rules. In the case of beta-decay, in which the transition takes place by the simultaneous emission of an electron and a neutrino, one may observe the relative probabilities of the various ways in which these particles share the available energy. This is done by observing the 'shape' of the beta-ray spectrum. While the theory (Konopinski 1943) is not on as firm a footing as that for electromagnetic transitions it seems to be adequate to interpret the available results in much the same manner as is the case for internal conversion ratios. A special case is the relative probability of positron emission and electron capture, similar in theory and experiment to the ratio of internal and pair conversion.

In nuclear reactions involving emission of heavy particles, the possibility of competing modes for transitions between the same states exists only in a few special cases.

Frequently the absolute transition probability is a clue to the nature of the states involved. For example, the cross section at the peak of a resonance for nuclear interaction with a neutron or gamma-ray is $\sigma = (\lambda^2/\pi)(2J+1)/(2I+1)$, where λ is the de Broglie wavelength of the incoming particle, J is the angular momentum of the compound state, and I that of the target nucleus. Thus the angular momenta can, to a certain extent, be derived from the cross sections. Similarly the lifetime of excited states against emission of electromagnetic radiation is related to the multipole character of the transition and the angular momenta of the initial and final states (Segrè and Helmholz 1949, Axel and Dancoff 1949). Although a precise calculation of the lifetime would involve detailed knowledge of nuclear wave functions, the dependence of the transition probability on angular momenta and parities of the states seems strong enough to permit significant conclusions to be drawn.

A similar situation exists for beta-ray transitions. The theoretical calculations are complicated by the fact that the interaction giving rise to the transition may contain a variety of terms analogous to the electric and magnetic terms in gamma-ray emission but rather more complex. Much new experimental material has become available in the last few years. In a considerable number of cases of so-called allowed transitions the shape of the beta-spectrum and the lifetime are in complete agreement with theory. The same is now true for several forbidden transitions (Shull and Feenberg 1949). In a few cases the spins of parent and product nuclei have been measured directly, and in all cases but one the agreement with beta-ray evidence is good. Although accumulated evidence has thrown light on the dominant terms in beta-interaction, some uncertainty remains. It seems probable that an important contribution arises in a manner analogous to the internal conversion of magnetic gamma-radiation (axial vector interaction), but it is not clear whether these are the only terms. Therefore it is not entirely safe to interpret the nature of nuclear states solely on the basis of beta-ray evidence.

For processes involving heavy particles, the transition probability does not depend strongly on the angular momenta of the nuclear states because of the shorter wavelengths involved. For example, in the case of alpha-radioactivity the effect of spin change is completely masked by small fluctuations in nuclear radius (Perlman, Ghiorso and Seaborg 1950).

Even when the transition probabilities themselves cannot be measured directly, interesting conclusions can often be drawn from relative intensities of transitions to two or more final states from the same initial level.

In atomic spectroscopy the angular momentum and other characteristics of a state are usually determined by Zeeman effect studies. This method does not seem at all feasible in the case of nuclear spectroscopy. Nuclear magnetic moments are of the order of 10^{-13} ev/gauss. The strongest magnetic fields now in sight could only produce a separation of the order of a few parts in 10^{11} for gamma-ray energies in the range of 10^5 ev. Even if gamma-ray energies could be measured with this accuracy, the natural width of most lines would mask any effect.

(iv) *Elementary Theory of Angular Correlations*

While it is thus impossible to investigate the individual Zeeman components of a transition by separating their energies, it is still possible to study their properties—spatial intensity distribution and polarization—if one can obtain a

non-uniform population of the magnetic substates of the initial state. This is the basis of all angular correlation experiments, at least in an elementary view. Fairly general theoretical discussions have been given, e.g. by Yang (1948) and by Falkoff and Uhlenbeck (1950 a). We shall illustrate the principles involved by a simple example and state some of the results. Consider a transition between two states of angular momentum quantum numbers $J=1$ and $J'=0$. In a 'weak' magnetic field this transition splits into a normal Zeeman triplet as shown in Figure 1 (*a*). In the absence of a magnetic field the total radiation from a sample is isotropic, since the atoms are oriented at random. The isotropy is preserved when the field is introduced. This is expressed in the principle of spectroscopic stability. The populations of the three magnetic substates with $m=+1, 0, -1$ are equal, as are the transition probabilities from the three states. This is a special case of general spectroscopic sum rules. The radiation from each of the sub-states, i.e. each of the Zeeman components, is not, in general, isotropic. Its angular distribution with respect to the direction of the magnetic field depends on the nature of the radiation emitted. For gamma-rays we have the well-known distributions for (electric or magnetic) dipole radiation,

$$F_1^{\,0}(\theta)=\sin^2\theta; \qquad F_1^{\,\pm 1}(\theta)=\tfrac{1}{2}(1+\cos^2\theta). \qquad \ldots\ldots(1)$$

The subscript indicates the multipole order, the superscript the change in magnetic quantum number. If the parity of the two states is the same, the radiation will be magnetic: if the parity differs it will be electric.

For alpha or other particles of zero spin the de Broglie wave of the outgoing particles must be a p-wave whose amplitude as a function of θ is given by an associated Legendre polynomial of order one and the intensity by its square. Thus in this case

$$F_1^{\,0}(\theta)=\cos^2\theta; \qquad F_1^{\,\pm 1}(\theta)=\tfrac{1}{2}\sin^2\theta. \qquad \ldots\ldots(2)$$

The parities of the two states must be different for the transition to be possible. The subscript to F now stands for the orbital angular momentum.

For deuterons or other particles of unit spin the situation is somewhat more complicated. It is seen that both equations (1) and (2) will give isotropic distributions if the three initial substates are equally populated, since $F_1^{\,0}+F_1^{\,+1}+F_1^{\,-1}=\sin^2\theta+\cos^2\theta=1$. However, if the populations are different, say $w(m)$ for the state of magnetic quantum number m, the angular distribution will be

$$W(\theta)=\Sigma w(m)F_1^{\,-m}(\theta)=w(0)F_1^{\,0}(\theta)+w(+1)F_1^{\,-1}(\theta)+w(-1)F_1^{\,+1}(\theta). \quad \ldots\ldots(3)$$

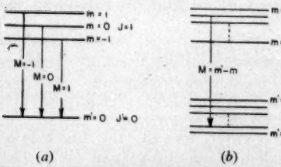
The most straightforward way to obtain non-uniform population $w(m)$ would be to apply a very strong magnetic field H at very low temperatures T. The populations would then be given by the Boltzmann factor $w(m)=e^{+\mu Hm/kTJ}$, where μ is the nuclear magnetic moment. Spiers (1948) has calculated the angular distributions in general cases for such experiments. Even with the strongest fields and lowest temperatures now available the effects expected are small, typical values being of the order of $0{\cdot}02(\mu H/kT)^2$. Preliminary results of Poppema (1950) are inconclusive. The nature of the experiments restricts them to rather few special cases.

In the more general situation illustrated in Figure 1 (*b*), the angular momentum J of the initial state is arbitrary and the latter splits into $(2J+1)$ m-states, with similar generality for the final state. The intensities of the several

components characterized by m and m' are no longer equal. If the emitted radiation contains a single L-value (multipole order or orbital angular momentum) the relative intensities are given by coefficients $G^{JLJ'}_{mm'}$, which depend only on the quantum numbers indicated but not on the nature of the radiation. The angular distribution of each component is $F^M_L(\theta)$, where $M=m'-m$. Then equation (3) is replaced by

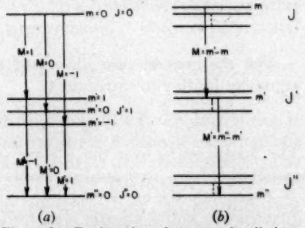
$$W(\theta)=\Sigma\Sigma w(m)G^{JLJ'}_{mm'}F^M_L(\theta). \qquad \ldots\ldots(4)$$

The summations extend over all values of m and m'. Only those coefficients G for which $m'-m=M$ is less than L are different from zero. If the $w(m)$ are known, equation (4) permits J, J' and L to be determined from the angular distribution, at least in principle. The $w(m)$ are usually known conditionally: for each assumed J they can be calculated. Because of the large number of possible combinations of parameters the answers can be assumed to be unique only if some other information limits the choice of angular momenta. This



(*a*) (*b*)

Figure 1. Designation of the magnetic substates for angular distributions.



(*a*) (*b*)

Figure 2. Designation of states and radiations for angular correlations.

is very often the case. It should be noted that equation (4) is equally applicable if the radiation is *absorbed* rather than emitted. Under the conditions for which equation (4) is valid $W(\theta)$ will be of the form

$$W(\theta)=\Sigma a_i\cos^{2i}\theta. \qquad \ldots\ldots(5)$$

The summation extends from $i=0$ to $i=J$ or $i=L$, whichever is smaller.

A method of more general applicability is to excite the nucleus to the initial state indicated in Figure 1 in such a way that the several m-states are formed with different probabilities. This method is illustrated in Figure 2. Figure 2 (*a*) shows a particularly simple example. The nucleus is initially in a state with $J=0$, and makes a transition to a state with $J'=1$, emitting or absorbing a radiation with $L=1$. This radiation is observed in the z-direction. This can be done either by choosing the direction of the bombarding beam in this direction in an experiment of the type described in § 3 (i) or by placing one counter along the z-axis in a correlation experiment (§ 3 (ii)). Actually the z-direction is arbitrary, since no magnetic field is used. The populations of the sub-states with $m'=1,\ 0,\ -1$ will then be given simply by

$$w(m')=F_1{}^{m'}(0), \qquad \ldots\ldots(6)$$

where the F may be those given in equation (1) or (2) or more complicated ones in other cases. Next it is assumed that the nucleus makes a further transition

to a state with $J''=0$, again emitting a radiation with $L'=1$, before perturbing fields have re-oriented it. This means that no transitions between the several m'-states take place within the lifetime of the intermediate state.

We can then use the result of (6) directly in (3) except for the change of m to m' and get

$$W(\theta)=F_1^0(0)F_1'^0(\theta)+F_1^{+1}(0)F_1'^{-1}(\theta)+F_1^{-1}(0)F_1'^{+1}(\theta). \quad \ldots\ldots(7)$$

The primed angular distribution functions refer to the second transition. For example, if both transitions are (dipole) gamma-rays, we use (1) in (7):

$$W(\theta)=0\times\sin^2\theta+1\times(1+\cos^2\theta)/2+1\times(1+\cos^2\theta)/2=1+\cos^2\theta. \quad \ldots\ldots(7')$$

This equation gives directly the angular correlation between two successive gamma-rays, if the angular momenta of the states are as indicated in Figure 2 (*a*). The first transition could also be an absorption process. In this case (7′) describes the angular distribution for resonance scattering of gamma-rays. If the first transition is an alpha-ray emission, and the second a gamma-ray, we find from (1), (2) and (7′)

$$W(\theta)=\sin^2(\theta). \quad \ldots\ldots(7'')$$

For the general case illustrated by Figure 8 (*b*) we find, by extending the argument leading to equation (4),

$$W(\theta)=\sum_{m'}\{[\sum_{m}G^{JLJ'}_{mm'}F^{M}_{L}(0)][\sum_{m''}G^{J'L'J''}_{m'm''}F'^{M'}_{L'}(\theta)]\}. \quad \ldots\ldots(8)$$

The double sums over the G's are somewhat tedious to carry out and general expressions have been published only up to $L=2$, $L'=2$.

Since only the angle θ between the two rays enters into (8), the direction of quantization could equally well have been taken along the second ray. The same result would be obtained. The same is true if the entire process illustrated in Figure 2 is reversed, i.e. if the final state replaces the initial one and vice versa. The following conditions must be satisfied to assure validity of equation (8). *a*. The magnetic substates characterized by m of the initial state must be equally populated. This is true if the nuclei are oriented at random but not necessarily true if the *initial* state is formed by bombardment with a beam of fast particles. *b*. The intermediate state must be a simple level with definite parity. This is not always true if the excitation is above the separation energy for a nucleon so that the states are wide and may overlap. *c*. The orientation of the nucleus must not be disturbed during the life of the intermediate state. This condition is usually satisfied for states of short lifetime since the interaction of nuclei with external fields is so weak. For metastable states this may not be true. Goertzel (1946) has discussed the perturbation to be expected from the magnetic field of the atomic electrons at the nucleus (hyperfine structure splitting). The condition for validity of (8) is that the reciprocal of the Larmor frequency in the perturbing field $\omega_L=g_N\mu_0H/\hbar$ should be large compared with the lifetime of the intermediate state. If this condition is not satisfied it may be possible to decouple the nucleus from atomic effects by an applied magnetic field in the direction of one of the radiations. Conversely, an external field perpendicular to the plane of the two rays should cause the correlation to precess with the Larmor frequency. It has been proposed by Sunyar, Alburger, Friedlander, Goldhaber and Scharff-Goldhaber (1950) to utilize this effect for the measurement of g_N, the nuclear Landé factor for states with measurable lifetime. *d*. Each of the two emitted or absorbed

radiations can be represented by a single partial wave or L-value. This condition is by no means generally satisfied and constitutes the most severe limitation on the applicability of (8). For example, in a gamma-ray transition between states with angular momenta J and J', all L values from $L=|J-J'|$ to $L=|J+J'|$ can contribute, alternate partial waves being electric or magnetic, depending on the parities of the states. It is generally assumed that at most two of these are of importance in any one transition; electric radiation with $L=|J'-J'|+1$ and magnetic radiation with $L=|J-J'|$ may interfere if the parity change is $(-1)^{|J-J'|+1}$. Otherwise only electric radiation with $L=|J-J'|$ appears. Except when J or J' is zero this possibility must always be considered. The same range of L from $J-J'$ to $J+J'$ can contribute in the case of heavy particles, but only odd *or* even L values will appear in transitions between states of definite parity. If the emitted particle has half-integer spin and relativistic velocity, even this restriction disappears, since parity in the usual sense is then no longer conserved (Yang 1948).

When the conditions for validity of equation (8) are satisfied, certain general statements can be made concerning the angular correlation. $W(\theta)$ is a polynomial in powers of $\cos^2\theta$,

$$W(\theta)=\Sigma a_i \cos^{2i}\theta, \qquad \ldots\ldots(9)$$

where i takes on all integral values from zero to $i=L$, L' or J', whichever is the smallest. Thus if the intermediate state has $J'=0$ or $J'=\frac{1}{2}$, the distribution is isotropic. For gamma-rays $L=0$ is excluded by the mechanism of radiation, but if in one of the transitions a particle is emitted or absorbed with zero orbital angular momentum (s-wave) an isotropic distribution results. Conversely, if a term in $\cos^{2i}\theta$ appears in the angular distribution or correlation, one may conclude that the intermediate state and each of the emitted or absorbed radiations has at least orbital angular momentum i. A detailed discussion of correlations of the type equation (9) is given by Falkoff and Uhlenbeck (1950 a).

Since $W(\theta)$ gives only the *relative* intensity in different directions it is customary to let $a_0=1$. Note that $W(\pi/2)=a_0$. Then

$$W(\theta)=1+a_1\cos^2\theta+a_2\cos^4\theta+\ldots \qquad \ldots\ldots(9')$$

Frequently a measure of the anisotropy is given in the form

$$[W(\pi)-W(\pi/2)]=(\Sigma a_i)-1,$$

but this parameter does not necessarily indicate the greatest deviation from isotropy since the a_i may have either sign. In § 3 (ii) we use $W(\theta)$ in a form normalized, so that $\int W(\theta)d\Omega=1$.

When condition *d.* above is not satisfied, the calculation is complicated by the fact that it is not the probabilities G and F appearing in equation (8) which must be combined, but the corresponding probability *amplitudes* with due regard to the (unknown) phases as well as magnitudes. The number of possibilities then becomes very large, but definite conclusions are still sometimes possible (Ling and Falkoff 1948, Spiers 1950). A similar situation arises when condition *b.* is violated (Devons and Hine 1949, Yang 1948).

We have assumed so far in this section that only the *direction* of the radiations is observed. It is also possible to observe their state of polarization. This has only been done for gamma-rays. Slow neutrons may lend themselves also to this type of observation, but it should be understood that significant effects can be expected

only if one has polarized nuclei available as well. The theory of the gamma-ray case has been discussed by Hamilton (1948) and Falkoff (1948). Its discussion would exceed the scope of this report. While the angular correlation, in principle, fixes the multipole order of the gamma rays L and L', the polarization correlation distinguishes between electric and magnetic multipoles of the same order and determines the parity change in the transitions.

The correlation functions (8) or (9) refer to the angles in the centre-of-mass system. If the nucleus acquires appreciable recoil volocity in the first transition the observed correlation in the laboratory coordinate system must first be transformed to the proper one before comparison with theory.

§3. EXPERIMENTAL TECHNIQUES

From the point of view of experimental technique, measurements of the intensity of nuclear radiation as a function of angle may be grouped into three main types: angular distributions, angular correlations, and experiments in which more than one angle is measured.

(i) *Angular Distributions*

The term angular distribution is usually applied to the relative emission probability as a function of the angle between the emitted ray or particle and a direction fixed by the experimental arrangement, for example the direction of an applied magnetic field or, more usually, of an incident beam of bombarding particles. A typical apparatus is shown in Figure 3. In this arrangement the number of scattered or secondary particles emitted from the small gas target into the solid angle defined by the counter slit is measured as a function of angle. A different arrangement, shown in Figure 4, uses an array of photographic plates to detect the secondary particles—in this case scattered protons—from a foil target.

Chadwick, May, Pickavance and Powell (1944) describe a similar method in which a single plate is used. A cloud chamber may be used in place of a photographic plate. The latter method was discussed in detail by Blackett and Lees (1932).

If the intensity of the emitted particles is recorded with a counter as in Figure 1, it is necessary to know the total bombardment accumulated by the target while data are taken at each angle. This problem is avoided in arrangements as in Figure 4, which record data at all angles simultaneously. It also does not arise if the bombarding particles are derived from a radioactive source whose strength is either practically constant or varies with time in a known way. Otherwise care must be taken that one measures the beam striking the effective target area, which may not be proportional to the total intensity delivered by the accelerating equipment. In Figure 3 the Faraday cage collects the beam after it has passed through the target.

It may also be important to control the bombarding energy quite accurately for several reasons. The reaction studied may show a resonance. A slight change in beam energy will then affect the yield quite sharply even if the intensity is kept constant. This would apply for example to the angular distribution of the gamma-rays from the reaction $^{19}\mathrm{F}(\mathrm{p},\alpha)\,^{16}\mathrm{O}^{*}$ (Van Allen and Smith 1941) at the very sharp resonance of 440 kev. (Bonner and Evans 1948). In this situation it is frequently preferable to monitor the emitted radiation at a fixed angle with a counter similar to the movable one instead of measuring the incident beam intensity.

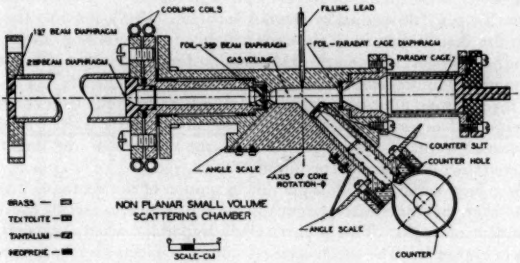


Figure 3. Gas target and counter arrangement used by Taschek (1948).

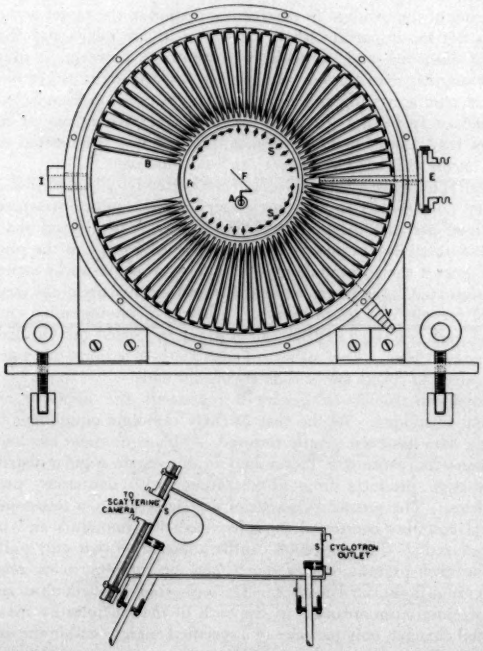


Figure 4. Scattering camera using sixty-six photographic plates according to Wilkins (1941). Protons enter at E and strike the target F. Slits S permit selection of different lengths of exposure at different angles.

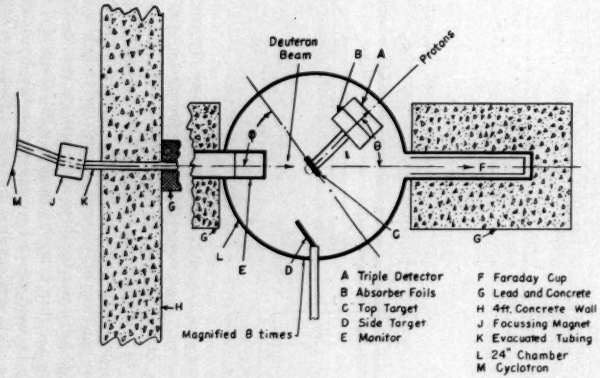
Such a device is illustrated by the counter M in Figure 6. Sometimes, as in the reaction $^{7}Li(p,\gamma)^{8}Be$ studied by Devons and Hine (1949), not only the yield but the angular distribution itself depends strongly on bombarding energy. In this case the necessity of close control of the bombarding energy cannot be avoided.

The main requirement for the target is that it be isotropic. The radiation under investigation must not be absorbed significantly in the target or its support on emerging at any angle at which measurements are to be taken. If an energy-dependent detector is used, energy loss of the radiation in the target must be considered since it will depend on the angle of emergence. The same is, of course, true for energy variation with angle due to motion of the centre of mass. Care must be taken that the effective target volume and beam area seen by the detector is independent of angle. This is particularly important when a gaseous target is used as in Figure 3. The effective target volume is selected by the intercept of the cone admitted by the counter collimator with the cylinder swept out by the beam, and the shape and location of this intercept varies with angle. This must be taken into account.

Scattering of the incident or emergent radiation in the target or surrounding material is not too important when both particles are heavy and charged. In the case of electrons, and particularly of neutrons, however, it may seriously distort the angular distribution. In the case of very light target nuclei or for electron–electron scattering this difficulty is sometimes overcome by detecting both secondary fragments in coincidence. Conservation laws of momentum and energy then assure that the process detected was the desired one. This method is hardly ever used in nuclear reactions, however.

It is desirable to use a detector of sufficient selectivity to discriminate against all particles not emitted in the particular reaction under investigation. For heavy charged particles the photographic plate or Wilson cloud chamber offer the simplest possibility of discrimination. For example, on the photographic plates in Figure 4 the point and angle of incidence of the tracks assure that the particles originated in the target, while the length and density of the tracks permit selection of protons of a given energy. The use of photographic emulsions in this way has been described in detail, for example by Powell and Occhialini (1947) and by Yagoda (1949). The analysis of such plates is tedious, and a great deal of work is required to collect statistically significant data.

The apparatus shown in Figure 5 represents the opposite extreme in experimental technique. At the cost of fairly elaborate equipment the labour of collecting data has been greatly reduced. This equipment has been used at the Massachusetts Institute of Technology to investigate angular distributions of charged reaction products from bombardment with deuterons, protons and alpha-particles. The secondary particles are detected in a telescope of three proportional counters connected to differential discriminators and to a triple coincidence circuit. These circuits can be adjusted so that only particles of a particular species, protons, deuterons, tritons, or alpha-particles, reaching the end of their range in the third counter will be registered. Selecting an appropriate thickness of aluminium absorber in the path of the particles by means of the automatic foil changer, only particles of a specified energy (within the uncertainty of range straggling) are counted. The beam intensity from the cyclotron is monitored by an ionization chamber placed in the incident beam from the cyclotron. The counter telescope collects counts until the monitor has collected



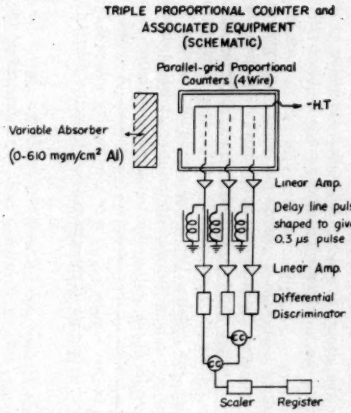


Figure 5. Target chamber and counter telescope used at Massachusetts Institute of Technology for the study of secondary charged particles.

a predetermined charge. Then the count is recorded, the absorber changed automatically and the apparatus reset. A detailed discussion of the evaluation of data obtained with this arrangement has been given by Gove (1950) and Boyer (1950).

When the emitted particle is a gamma-ray, careful energy selection is not usually made in angular distribution experiments. Magnetic spectrometers capable of selecting single gamma-ray energies are somewhat bulky for large angular motion, and their efficiency is low. Recent developments in scintillation spectrometers (Bell and Jordan 1949) will no doubt improve this situation. When the gamma-rays whose angular distribution is being studied are more energetic than the bulk of the undesirable background radiation, the latter can be eliminated by absorption of the secondary electrons. A typical arrangement of this type is shown in Figure 6. It was used by Devons and Hine (1949) to study the angular distribution of the gamma-rays from the reaction $^{7}Li(p,\gamma)^{8}Be$,

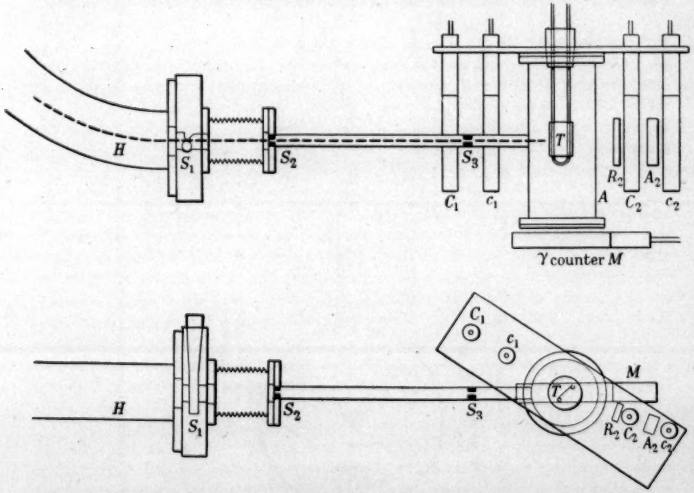


Figure 6. Arrangement of Devons and Hine (1949) for angular distributions of gamma-rays.

which have energies of 17·2 Mev. and 14·4 Mev., much higher than other radiations likely to be present. Secondary electrons produced by the gamma-rays in the radiator R are counted if they penetrate both Geiger–Müller counters C which are connected to a coincidence circuit. The absorber A placed between the counters stops secondary electrons from less energetic gamma-rays.

The selective detection of neutrons is particularly troublesome. Fast neutrons may be detected by first slowing them to thermal energies in hydrogenous material, such as paraffin, and then detecting them in a BF_3 counter. Instead of a counter a material may be used which becomes radioactive by neutron capture. This method may be made very sensitive by using a Szillard–Chalmers process. For

example, if the detector consists of a considerable volume of a $Ca(MnO_4)_2$ solution a very large fraction of the incident neutrons will be absorbed, forming radioactive ^{56}Mn which may be separated from the solution as MnO_2 by filtration (Meiners 1949). This technique, as well as the Boron counter, is unsuitable for selecting definite neutron energies. The sensitivity of these detectors does depend somewhat on the energy, however. This may be troublesome in reactions involving light nuclei where the neutron energy from the reaction depends strongly on angle because of the centre-of-mass motion. Fast neutrons are also detected through recoil protons. These may be produced in a thin film of hydrogenous material such as polyethylene and then detected in a proportional counter or a photographic plate. In the latter case the gelatine of the emulsion itself may be used as 'converter'. The energy of the protons can be measured by the methods used for charged particles originating in the target itself. The neutron energy is determined only if the angle between incident neutron and recoil proton is also known.

Counters filled with a hydrogen-containing gas, such as methane, at a high pressure have a higher efficiency than those with a solid converter because the thickness of the latter must be kept small compared with the range of the recoil protons if any energy discrimination is desired. The pulse–height distribution from a methane-filled proportional counter represents the energy spectrum of the recoil protons directly. A beam of mono-energetic neutrons will, however, produce a continuous proton spectrum, uniform in intensity from zero energy to the full neutron energy. This device can therefore be used to select distinct groups of neutrons only if the spectrum is fairly simple. The problem of scattering from surrounding materials is more serious for neutrons than for other radiations. Angular distributions of neutrons are often distorted by this 'room scattering'.

It is frequently necessary to correct the observed distribution for the effect of finite angular resolution of the detector. If the 'line profile' of the arrangement is symmetrical and independent of angular position this correction is very simple. The condition may be stated in the following form: The probability of detecting a particle emitted at an angle θ when the midpoint of the detector is located at an angle θ_0 is given by a function $f(\alpha)$ of α only, where $\alpha=(\theta_0-\theta)$ and $f(-\alpha)=f(\alpha)$. Then the observed intensity $W'(\theta_0)$ is related to the true intensity $W(\theta_0)$ by the relation

$$W'(\theta_0)=W(\theta_0)+\frac{1}{2}\frac{d^2}{d\theta^2}W(\theta_0)\frac{\int\alpha^2 f(\alpha)\,d\alpha}{\int f(\alpha)\,d\alpha}+\frac{1}{4!}\frac{d^4}{d\theta^4}W(\theta_0)\frac{\int\alpha^4 f(\alpha)\,d\alpha}{\int f(\alpha)\,d\alpha}+\ldots.$$

where the integrations extend over the full angular spread of the detector. The correction is thus essentially proportional to the second derivative of $W(\theta)$. In many practical cases $f(\alpha)$ is also a function of θ_0. This is always true when detector and source have a finite extent perpendicular to the plane of θ. A detailed analysis of the geometric conditions is then required.

For comparison with theoretical calculations the angular distributions must be given in the centre-of-mass coordinate system for the process. In the non-relativistic form applicable to most heavy-particle reactions the angle θ' in this coordinate system is expressed in terms of the angle θ in the laboratory system by $\sin(\theta'-\theta)/\sin\theta=v_{c.g.}/v'$, where $v_{c.g.}$ is the relative velocity of the two systems and v' is the velocity of the emitted particle in the centre-of-mass system. If the

emitted particle is a gamma-ray the expression becomes $\cos\theta(1+\beta\cos\theta)=\cos\theta'+\beta$, where $\beta=v_{c.g.}/\boldsymbol{c}$. Not only the angle but also the solid angle must be transformed to the new coordinate system by multiplication by $f(\theta)=(\sin\theta'\,d\theta')/(\sin\theta\,d\theta)$.

(ii) *Angular Correlations*

The (conditional) probability that one of a pair of rays be emitted along a given direction into an element of solid angle $d\Omega$ if the other is emitted at an angle θ to it is the angular correlation of the two rays. The experimental problem which distinguishes correlation experiments from measurement of distributions is the necessity to identify two rays as the members of a correlated pair, emitted by the same atom. If both rays are charged particles their tracks may be seen to originate from the same point in a cloud chamber or photographic plate, and the angle between them may be measured. Thus the correlation between the alpha-particles from the break-up of ^{8}Be following beta-decay of ^{8}Li was measured in a cloud chamber (Christy, Cohen, Fowler, Lauritsen and Lauritsen 1947). The limitations on this technique are essentially the same in the study of

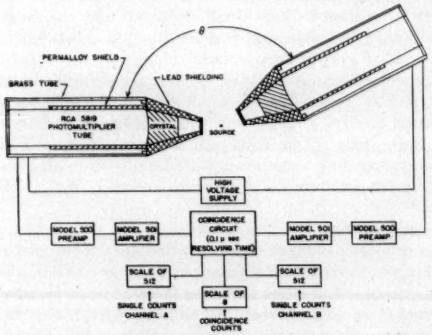


Figure 7. Apparatus for measuring angular correlations of gamma-rays.

correlations as those discussed under § 3 (i). Therefore, when good statistical accuracy is required, or whenever one of the rays is a gamma quantum, one uses coincidence in *time* rather than in space to identify a pair of rays originating from the same atom. General features of the coincidence method have been discussed by a number of authors (e.g. Feather and Dunworth 1938, Dunworth 1940, Maier-Leibnitz 1946). Its outstanding characteristic is the fact that the rate of data-taking, and therefore the statistical accuracy attained, is limited by properties of the detector and electronic circuits, which cannot be compensated by increasing the strength of the source. A typical arrangement for an angular correlation coincidence experiment is shown in Figure 7.

Consider a source which emits successively two radiations, 1 and 2, e.g. an alpha-particle and a gamma-ray. Counter A might be a proportional counter designed to detect only the alpha-particles with an efficiency $\epsilon_A(1)$, subtending a solid angle Ω_A at the source. Counter B detects only gamma-rays with efficiency $\epsilon_B(2)$ and solid angle Ω_B. It might be an appropriate scintillation

counter. The source undergoes N_0 disintegrations per second, emitting a pair of rays in each. The angular correlation between the alpha- and the gamma-rays $W(\theta)$ is assumed to be a sufficiently slowly varying function of θ that its value at the angle joining the midpoints of the counters may be used in first approximation. Then the observed 'true' coincidence rate $N(\mathrm{AB})$ will be

$$N(\mathrm{AB}) = N_0 W(\theta) \epsilon_{\mathrm{A}}(1) \epsilon_{\mathrm{B}}(2) \Omega_{\mathrm{A}} \Omega_{\mathrm{B}} / 4\pi^2. \qquad \ldots\ldots(10)$$

Corrections can be made for the finite angular resolution of the counters as in the case of angular distributions. Because of the finite spread for both rays this can become an awkward problem in solid geometry, particularly if the efficiency of the counters is not uniform over their faces.

In practice simultaneity of the two rays can be ascertained only within the finite resolving time of the circuit used. Thus if $\tau(\mathrm{AB})$ is the longest interval by which a count in B may succeed one in A and still be registered as simultaneous, while $\tau(\mathrm{BA})$ is the corresponding time for the reverse order of occurrence, the 'resolving time' τ is usually defined by the relation $\tau = \frac{1}{2}[\tau(\mathrm{AB}) + \tau(\mathrm{BA})]$. If the counting rates in the two counters $N(\mathrm{A})$ and $N(\mathrm{B})$ are small compared with the coincidence rate $N(\mathrm{AB})$, the rate of accidental coincidences due to counts arising from unrelated events is, to a good approximation (Dunworth 1940),

$$N_{\mathrm{acc}} = 2\tau N(\mathrm{A}) N(\mathrm{B}). \qquad \ldots\ldots(11)$$

With the same assumptions about decay scheme and efficiencies as were used to derive (10), and the additional condition that background counting rates are negligible, (11) becomes

$$N_{\mathrm{acc}} = 2\tau N_0^2 \epsilon_{\mathrm{A}}(1) \epsilon_{\mathrm{B}}(2) \Omega_{\mathrm{A}} \Omega_{\mathrm{B}} / (4\pi)^2. \qquad \ldots\ldots(11')$$

The accidental coincidence rate rises with the square of the source strength while the true rate rises only linearly. Therefore the rate of accumulating statistically significant data is not raised appreciably by increasing the source strength beyond the value $N_0 = 1/2\tau$, for which the true and accidental coincidence rates are equal. For this particular source strength the true rate becomes

$$N(\mathrm{AB}) = W(\theta) \epsilon_{\mathrm{A}}(1) \epsilon_{\mathrm{B}}(2) \Omega_{\mathrm{A}} \Omega_{\mathrm{B}} / 2\tau(4\pi)^2. \qquad \ldots\ldots(12)$$

In most experiments the lower limit of τ is determined by the fluctuations in response time of the counters. For ionization detectors, including Geiger–Müller counters, this limit is about 2×10^{-7} sec. With scintillation counters this can be improved by about a factor of ten. Other considerations of electronic circuit design may set a higher limit. In some cases the time interval between the emission of the two rays may itself set a limit.

The angular spread $\delta\theta$ of the radiation admitted to the counters limits the resolution of the resulting curve. Having chosen an optimum shape for the detectors, it is usually best to make the spread equal for A and B. Then $\Omega_{\mathrm{A}} = \Omega_{\mathrm{B}} \sim (\delta\theta)^2$ and $N(\mathrm{AB}) \sim (\delta\theta)^4$, i.e. the rate of obtaining data is inversely proportional to the fourth power of the angular resolution. Therefore good resolution can be had only at a great sacrifice in counting rate. If $W(\theta)$ is a slowly varying function of θ, as was assumed in deriving equation (10), angular spreads of $\pm 15°$ are usually quite admissible. The effect of this spread on the correlation can be corrected for by calculation. When $W(\theta)$ is a very sharp function of θ equation (12) is not applicable.*

* For example, in the case of the annihilation of slow positrons the two gamma-ray quanta are emitted practically exactly in opposite directions, so that $W(\theta)$ is practically a δ-function. In this case, for $\theta \neq \pi$, $N(\mathrm{AB}) = 0$, and for $\theta = \pi$, $N(\mathrm{AB}) \simeq \epsilon_{\mathrm{A}} \epsilon_{\mathrm{B}} / 2\tau$ independent of Ω, if the source strength can be suitably chosen.

Equation (12) also illustrates the great importance of achieving a high efficiency for the detectors A and B. While this efficiency is usually near unity for corpuscular radiations, it can be quite low for non-ionizing rays, i.e. neutrons and gamma-rays. For example, in the case of gamma–gamma angular correlations the older experiments (Kikuchi, Watase and Itoh 1942, Beringer 1943, Good 1946) used Geiger–Müller counters as detectors. These have efficiencies of the order of one per cent for gamma-rays. As a result $N(\mathrm{AB})$ had values of the order of 0·1 count per second under typical conditions. The introduction of scintillation counters to this type of experiment (Brady and Deutsch 1950) raised the efficiency by more than a factor of ten. Therefore the useful counting rate was raised one hundred-fold.

When it is impossible to use counters sensitive only to the relevant radiations 1 and 2 respectively, the quality of the data suffers, as in the case of angular distributions discussed under § 3 (i). The effect is somewhat more complex in correlation experiments. In a simple experiment in which only one angle enters, i.e. when there is no fixed external direction or sense of rotation, $W(-\theta)=W(\theta)$. Then the positions of the two counters are equivalent, and it does not matter which of the two radiations is recorded in each direction. The product $\epsilon_{\mathrm{A}}(1)\epsilon_{\mathrm{B}}(2)$ in equation (10) is then replaced by $[\epsilon_{\mathrm{A}}(1)\epsilon_{\mathrm{B}}(2)+\epsilon_{\mathrm{A}}(2)\epsilon_{\mathrm{B}}(1)]$, and in equation (11′) by $[\epsilon_{\mathrm{A}}(1)+\epsilon_{\mathrm{A}}(2)][\epsilon_{\mathrm{B}}(1)+\epsilon_{\mathrm{B}}(2)]$. Equation (12) becomes

$$N(\mathrm{AB})=\frac{W(\theta)}{2\tau}\frac{\Omega_{\mathrm{A}}\Omega_{\mathrm{B}}}{(4\pi)^2}\frac{[\epsilon_{\mathrm{A}}(1)\epsilon_{\mathrm{B}}(2)+\epsilon_{\mathrm{A}}(2)\epsilon_{\mathrm{B}}(1)]^2}{[\epsilon_{\mathrm{A}}(1)+\epsilon_{\mathrm{A}}(2)][\epsilon_{\mathrm{B}}(1)+\epsilon_{\mathrm{B}}(2)]}. \quad \ldots\ldots(13)$$

If there are other radiations, from the source under investigation or of other origin, which cause counts in A and B but no true coincidences, the accidental coincidence rate is correspondingly raised. If only a fraction $f(\mathrm{A})$ of the counts in A is due to the desired radiations and correspondingly $f(\mathrm{B})$ in B, the right-hand side of equations (12) and (13) must be multiplied by the product $f(\mathrm{A})f(\mathrm{B})$. If other radiations also cause coincidences their effect must be investigated in a separate experiment. It is of course always desirable to use a detector which selects only the radiations under study, but sometimes this is possible only by sacrificing overall efficiency. A compromise must then be made.

The requirement that the source arrangement itself be isotropic in correlation experiments is similar to that in distributions. In practice the condition is somewhat less stringent for correlations. In a typical experiment, A might remain stationary with respect to the source while B is rotated. If the source is anisotropic, $\epsilon_{\mathrm{B}}(2)$ in equation (10), and therefore $N(\mathrm{AB})$, may depend on angle. However, the counting rate in B, $N(\mathrm{B})=N_0\Omega_{\mathrm{B}}\epsilon_{\mathrm{B}}$, depends on the anisotropy of the source in the same manner as $N(\mathrm{AB})$, and we find

$$N(\mathrm{AB})/N(\mathrm{B})=\text{const.}\times W(\theta) \quad \ldots\ldots(14)$$

even if the source is not located symmetrically or if there is a variation of absorption in different directions. Equation (14) is not necessarily valid if the radiations are more complicated than is implied by equation (10). It is essential, as in all quantitative coincidence experiments, that any effects of source structure have no correlation between the two counters. This means in practice that the extent of the source should be small compared with the distance from the counters and that self-absorption for *one* of the radiations should be negligible. The effect of scattering by the source and by surrounding material is the same as for distribution experiments discussed in § 3 (i).

A very insidious source of error arises from the possibility that a single ray from the source may, by scattering or other secondary processes, actuate both counters A and B. This effect, which is especially marked when both counters detect either electrons or gamma-rays, usually depends strongly on the relative positions of the counters. Figure 8 shows a result for gamma–gamma coincidences (Beringer 1943) in ^{64}Cu. The rise for $\theta<90°$ is entirely spurious, due to Compton scattering events in which the recoil electron is recorded in one counter and the scattered quantum in the other. The arrangement in Figure 7 is designed to eliminate this difficulty. The thick lead walls of the shielding cones around the detecting crystals of the scintillation counters absorb scattered

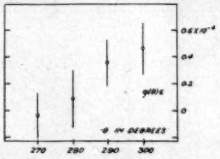


Figure 8. Coincidence rate of unshielded counters as a function of angular position according to Beringer (1943). The effect is due to Compton scattering.

quanta when θ is small. The extension of the cones beyond the end of the crystals to a small opening serves to reduce the effect near $\theta=180°$. Since the energy of back-scattered quanta is always less than 250 kev., the thin lead absorbers over the openings serve to reduce the effect of scattering when the two counters face each other. A new arrangement of this sort should always be tested by investigating the coincidence rate, as a function of angle, obtained with a source emitting only non-coincident radiations of nearly the same energy as are expected in the actual experiment. Any coincidences observed above the accidental rate can then be minimized by suitable shielding or corrected for in the results.

(iii) *Experiments involving more than one Angle*

Sometimes angular correlations between two rays are measured in an experiment where some other direction has a significant property. An example of this is the correlation between the alpha-particles and gamma-rays from the reaction ^{19}F (p, α) ^{16}O (Arnold 1950), in which the direction of the incoming proton is also known. Here no new experimental features are introduced compared with simple angular correlation experiments, although the results may differ considerably from those obtained when the incoming radiation is isotropic.

Another example is the measurement of the correlation between angle of emission and plane of polarization for two successive gamma-rays. Figure 9 illustrates the arrangement used by Metzger and Deutsch (1950) for this purpose. The direction of polarization of one of the gamma-rays is determined statistically by observing the direction of the scattered quantum when the ray undergoes a Compton encounter in the scattering crystal B. Scattering is more probable in a direction perpendicular to the direction of the electric vector. At the same time the direction of the other gamma-ray is determined by counter A. Thus two angles enter the experiment: angle θ between the directions of emission of

the two quanta and angle ϕ between the scattered quantum and the plane defined by the two primary quanta. Under practical experimental conditions the counting rate $N(C)$ in counter C and the coincidence rate $N(AC)$ are predominantly due to rays striking C directly from the source, rather than scattered quanta from B. Therefore B itself was made to be a scintillating crystal and its pulses

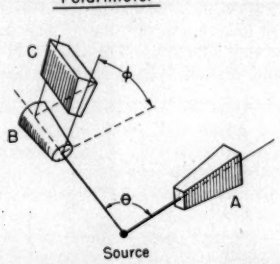


Figure 9. Schematic view of the gamma-ray polarimeter of Metzger and Deutsch (1950). A, B and C are crystals of fluorescent material whose light output is recorded by photocells (not shown).

observed in coincidence with A and C. The complications arising from asymmetries and scattering are rather more severe in this experiment than in a simple correlation experiment. They are discussed in detail in the paper quoted (Metzger and Deutsch 1950). The limiting useful source strength is again of the order $N_0 \simeq 1/2\tau$.

§ 4. EXPERIMENTAL RESULTS

(i) *Induced Reactions*

A very complete bibliography of reactions of light nuclei up to ^{20}Ne, including references to angular distribution experiments, can be found in a recent review article by Hornyak, Lauritsen, Morrison and Fowler (1950). We select here a few of the more readily interpreted interesting results only.

The angular distribution of the gamma-rays with respect to the incoming protons in the reaction $^3H(p,\gamma)^4He$ has been investigated (Argo, Gittings, Hemmendinger, Jarvis and Taschek 1950) at two proton energies. Corrected for motion of the centre of mass the results are: at $E_p = 1{\cdot}52$ MeV., $W(\theta) = \sin^2\theta + 0{\cdot}11$, at $E_p = 2{\cdot}16$ MeV., $W(\theta) = \sin^2\theta + 0{\cdot}09$. These results are so close to pure $\sin^2\theta$ distributions that they suggest the following interpretation: The spins of the proton and triton are anti-parallel in the reaction and the proton enters predominantly as a p-wave, forming a virtual excited $^1P_1$ state of the alpha-particle which decays to the $^1S_0$ ground state with the emission of electric dipole radiation. This is exactly the situation described in Figure 2(*a*) and equation (7^r). The slight deviation from a pure $\sin^2\theta$ distribution shows that other states must also contribute to the reaction. The excitation energy of the virtual state of the alpha-particle involved is about 24 MeV., in the same region in which strong dipole transitions are observed in heavier nuclei. The cross section of the reaction is so large that it indicates an unusually great dipole strength.

A similar result, namely an almost pure $\sin^2\theta$ distribution, has been found by Fowler, Lauritsen and Tollestrup (1949) in the reaction $^2H(p,\gamma)^3He$.

In the reaction $^2H(\gamma,p)^1n$ two processes may occur at moderate energies. One is essentially the reverse of the situation just discussed: an electric dipole quantum is absorbed and raises the nucleus from the $^3S_1$ ground state to a ^3P state. This should again give rise to a $\sin^2\theta$ distribution. The other process is one in which a ***magnetic*** dipole quantum re-orients the relative spin orientation of the neutron and proton, forming a $^1S_0$ state. Since the compound state has zero angular momentum, an isotropic distribution must result. Normally the electric dipole interaction should be expected to be rather stronger than magnetic dipole interaction. The low-lying virtual $^1S_0$ state of the deuteron provides a broad resonance for the magnetic process, however, so that for gamma-ray energies only slightly above the threshold the isotropic term predominates. The theory of this process is discussed in the review article by Bethe and Bacher (1936) and has since been extended by Rarita and Schwinger (1941). If we denote the ratio of the electric to the magnetic cross section by R, the angular distribution should be $W(\theta)=1+\frac{2}{3}R\sin^2\theta$. The factor $\frac{2}{3}$ arises from normalization of the $\sin^2\theta$ function. Woodward and Halpern (1949) found R to be 1·5 for a gamma-ray energy about 0·3 MEV. above threshold. At higher energies R increases and reaches about 4 for gamma-rays 0·66 MEV. above threshold. A number of other papers on this reaction have appeared in recent years (Graham and Halban 1945, Genevese 1949, Lassen 1949, Meiners 1949).

The reaction $^2H(d,p)^3H$ has been the subject of numerous investigations. An extensive bibliography of the more recent work can be found in the review paper by Hornyak *et al.* (1950). The angular distribution of the protons is of the form $W(\theta)=1+a_1\cos^2\theta+a_2\cos^4\theta$. a_1 rises from a value of 0·18 at the very low deuteron energies to 1·5 for 0·39 MEV. deuterons. At 1·5 MEV. it falls to zero and becomes negative at higher energies. a_2 appears first for deuterons above 1 MEV. It is positive and increases with increasing energy. The theory of this reaction has been discussed by Beiduk, Pruett and Konopinski (1950).

In a considerable number of reactions of light nuclei the distributions are isotropic, for example $^3H(p,n)^3He$ for proton energies between 1·2 and 2 MEV. (Taschek 1949), $^6Li(p,\alpha)^3He$ at 0·2 MEV. (Neuert 1937), $^7Li(d,n)^8Be$ at 0·605 MEV., 0·700 MEV., 0·820 MEV., and 1·34 MEV. (but not at 1·02 MEV.), where the neutrons are predominantly forward (Bennett, Bonner, Richards and Watt 1947, Whaling, Evans and Bonner 1949), $^9Be(d,\alpha)^7Li$ at 0·7 MEV. (Resnick and Hanna 1949), $^{10}B(n,\alpha)^7Li$ for slow neutrons (Gilbert 1948), $^{11}B(p,\alpha)$ at 0·16 MEV. and 0·20 MEV. (Jacobs and Wilson 1941), $^{12}C(p,\gamma)^{13}N$ at 0·5 MEV., $^{13}C(p,\gamma)$ at 0·625 MEV. (Devons and Hine 1949), $^{15}N(d,\alpha)^{12}C$ at 0·34 MEV. and 0·40 MEV. (Cochrane and Hester 1949). In view of the relatively low bombarding energies used it is perhaps not surprising to find formation of the compound nucleus by an s-wave frequently preponderant. This process will always give rise to isotropic distributions, whatever the other angular momenta involved. In the reaction $^9Be(p,\alpha\gamma)^6Li$ the angular distribution of the gamma-rays is isotropic at 340 KEV. (Jacobs, Malmberg and Wahl 1948), indicating proton absorption with $L=0$, but at 988 KEV. the distribution is $W(\theta)=1+0{\cdot}09\cos^2\theta$ (Devons and Hine 1949), interpreted as a small contribution from d-wave protons with $L=2$. The reaction $^{19}F(\alpha,p)^{22}Ne$ at 3·5 MEV. gives rise to several short-range groups of protons leaving the neon nucleus in excited states, besides the group leading to

the ground state. The short-range protons are emitted isotropically while the ground state group shows a higher intensity in the forward direction. This indicates that both odd and even L-values of the alpha-particles must contribute to the formation of the compound nucleus since odd powers of $\cos\theta$ are required to give the asymmetry between forward and backward directions, violating equation (9) (Champion and Roy 1948).

When the angular distribution is observed at a resonance for a reaction a single state of the compound nucleus accounts for practically all of the reaction. Condition b of §2(iv) is then satisfied and the distribution should have the form

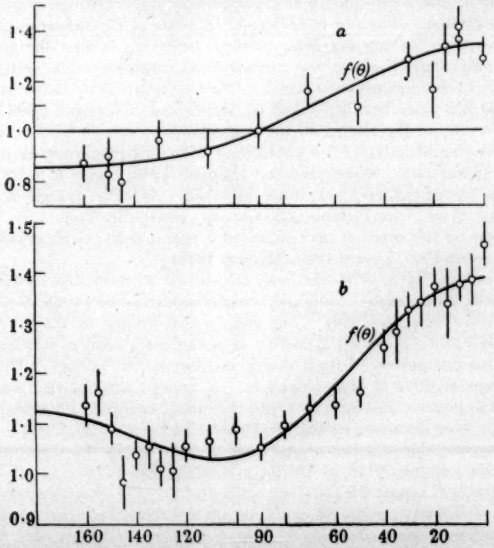


Figure 10. Angular distribution of gamma-rays from ^{8}Li (p, γ) ^{8}Be according to Devons and Hine (1949). *a*. Proton energy 0·650 MeV., $f(\theta)=1+0{\cdot}25\cos\theta+0{\cdot}1\cos^2\theta$; *b*. Proton energy 0·890 MeV., $f(\theta)=1+0{\cdot}12\cos\theta+0{\cdot}20\cos^2\theta$.

of equation (9), although condition d will generally not be satisfied, so that equation (8) is not applicable. In a non-resonant reaction states of opposite parities may overlap in the compound nucleus, equation (9) no longer holds, both odd and even powers of $\cos\theta$ may appear in the distribution, and the interpretation can become very difficult.

A striking example of this is provided by the distribution of the gamma-rays from the reaction ^{7}Li (p, γ) ^{8}Be (Devons and Hine 1949), in the neighbourhood of the 440 kev. resonance. At the resonance energy the distribution is very nearly isotropic, indicating that the protons enter with zero angular momentum. Above and below the resonance energy the distribution follows the form $W(\theta)=1+a\cos\theta+b\cos^2\theta$, as illustrated in Figure 10. The appearance of an

odd power of the cosine can only be explained by the interference of two states in the compound nucleus having opposite parity. The variation of this odd-power term as a function of energy as measured by the ratio $W(15°)/W(145°)$ is shown in Figure 11. Presumably the contribution of p-wave protons is important at all energies other than the resonance. The situation is complicated somewhat by the appearance of two different gamma-ray energies whose ratio varies with bombarding energy.

A very extreme case is observed for the protons from the reaction ^{6}Li (d, p) ^{7}Li whose angular distribution has been observed for deuteron energies from 0·4 to 1·4 Mev. (Whaling and Bonner 1950). Both odd and even powers of $\cos\theta$ up to the *sixth* appear, particularly at the higher energies. The distributions show a preference for emission in the forward direction. On the other hand the alpha-particles from the reaction ^{6}Li (d, α) α and the very similar ^{7}Li (p, α)α (Inglis 1948, Resnick and Hanna 1949) show distributions of the form

$$W(\theta) = 1 + a_1 \cos^2\theta + a_2 \cos^4\theta,$$

although the magnitude of the coefficients indicates that protons with $L = 3$ also contribute to the second of these reactions. In this case, where the reaction

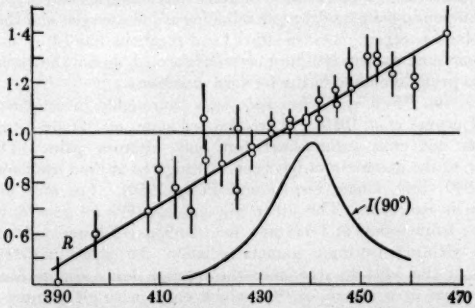


Figure 11. Asymmetry of the gamma-rays from ^{7}Li (p, γ) ^{8}Be as a function of bombarding energy near the 440 kev. resonance, according to Devons and Hine (1949). Curve I (90°) shows the gamma-ray yield at right angles to the beam for energies near resonance.

results in the production of two identical particles of zero spin, additional limitations are placed on the outgoing wave functions by symmetry considerations.

There is strong evidence that proton distributions from (d, p) reactions are not well described in terms of states of a compound nucleus, particularly at higher energies. For example, with deuterons of 14·6 Mev. (Gove 1950) the resulting protons always go predominantly forward, and except in a few cases the distributions seem to depend only on the approximate mass of the target nucleus and the energy release, but not markedly on the individual reaction. The intensity is about twenty times as high at 20° as at 140°. For the heaviest target nuclei, for example gold, the intensity has a maximum at about 60°. This behaviour can be interpreted by assuming that the deuteron breaks up in the field of the nucleus, only the neutron entering, while the proton is repelled by the electrostatic force. To a certain extent the reverse process, in which the proton enters the nucleus while the neutron remains outside, also seems to happen at

these energies (Falk, Creutz and Seitz 1949). Indeed, at 200 MEV. the neutrons from (d, n) reactions are concentrated in a narrow forward cone and their energy and angular distribution is well explained by a 'stripping' of the deuteron at the nuclear surface (Serber 1947).

At lower energies an intermediate situation seems to exist in the reaction $^{16}O(d, p)^{17}O$ at energies below 3 MEV. as investigated by Heydenburg and Inglis (1948), and perhaps in the reaction $^{27}Al(d, p)^{28}Al$ studied by Holt and Young (1950) and by Gove (1950). The latter case is illustrated in Figure 13. In the process $^{12}C(d, p)^{13}C$, which shows definite resonances, the angular distribution is essentially symmetrical about 90°, as one should expect for a resonance reaction (Heydenburg, Inglis, Whitehead and Hafner 1949).

Some other deuteron reactions also show consistently greater forward intensity of the secondary particles, indicating persistence of momentum rather than compound nucleus formation. At 14·6 MEV. the tritons and the inelastically scattered deuterons from several targets throughout the periodic table show this behaviour (Boyer 1950), and the distribution of tritons produced from beryllium by deuterons is well interpreted by a 'pick-up' process. All of these reactions must be considered to be intermediate between the cases discussed in § 2 (ii) and (iii), with the compound-state model predominating at low energies and the scattering picture at high energies. On the other hand reactions like (d, α) and (α, n), in which a compound nucleus is almost certainly formed, do not show a very striking tendency to prefer emission in the forward direction.

The reaction $^{19}F(p, \alpha)^{16}O$ has been very thoroughly investigated by many methods (Hornyak *et al.* 1950). Three excited states of ^{16}O are formed, one of which does not emit gamma-rays but only electron pairs. The angular distribution of the gamma-rays has been investigated at nine resonance energies (Devons 1949, Day, Chao, Fowler and Perry 1950). For four of these the distribution is isotropic. The other five yield $W(\theta) = 1 + a\cos^2\theta$, with values of a ranging from $-0{\cdot}08$ at 1·381 MEV. to $+0{\cdot}69$ at 1·290 MEV. At three of the resonances yielding isotropic gamma-radiation the alpha-particles are also isotropic. At the 340 KEV. resonance the reaction leads almost exclusively to an excited state at 6·15 MEV. in ^{16}O, which emits a single gamma-ray in the transition to the ground state. Arnold (1950) studied the angular correlation between the alpha-particles and gamma-rays for this resonance and found $W(\theta) = 1 + 111\cos^2\theta - 305\cos^4\theta + 225\cos^6\theta$. Since the two radiations individually are isotropic with respect to the protons, the magnetic substates of the ^{20}Ne compound nucleus will have 'natural' populations when quantized in the direction of emission of either of the rays. Also, since the ground state of ^{16}O has zero angular momentum, only one L-value can contribute to the gamma-radiation. Therefore equations (8) and (9) are applicable. The observed correlation is explained by assuming $J = 1$ for the 340 KEV. resonant state of ^{20}Ne, with even parity, $J' = 3$, odd parity for the excited state of ^{16}O and $J'' = 0$, even, for the ground state. The alpha-particles then carry $L = 3$ and the gamma-rays are electric octupole radiation. This description of the state in neon explains the absence of direct alpha-transitions to the ground state of oxygen since the p-wave required for the alpha-particles by the angular momenta would involve parity change.

Another state in ^{16}O which is also formed in this reaction, but at different resonances, has an excitation of 6·04 MEV. and decays only by emission of electron

pairs. This can be understood on the assumption that it has zero angular momentum and the same parity as the ground state (Oppenheimer and Schwinger 1939). The angular correlation between the electron and positron and the energy distribution of the particles should then be (Oppenheimer 1941)

$$P(W_+, \theta)\, dW\, d\theta = \text{const.} \times p_+ p_- (W_+ W_- + p_+ p_- \cos\theta - m_0^2 c^4)\, dW\, d\theta, \quad \ldots\ldots (15)$$

where p is the momentum and W the total energy of the particle in relativistic units. Devons and Lindsey (1949) measured this correlation and found $W(\theta) = 1 + 0{\cdot}85 \cos\theta$, in good agreement with theory. The energy distribution of the positrons also agrees well with expectation (Kojima 1943, Rasmussen, Hornyak, Lauritsen and Lauritsen 1950). Finally the lifetime of the state, 7×10^{-11} sec. (Devons, Hereward and Lindsey 1949), is consistent with the same interpretation.

Thus the angular momenta and parities of two excited states in ^{16}O besides the ground state are now well established, largely on the basis of angular correlation experiments. The energy level scheme is shown in Figure 12.

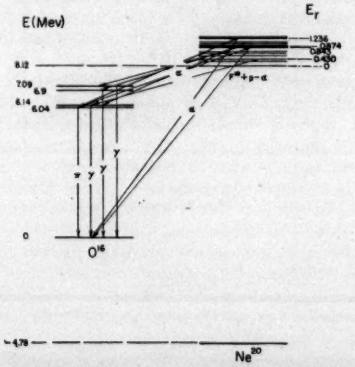


Figure 12. Some energy levels of ^{16}O and ^{20}Ne. The energy scale on the left refers to the ground state of ^{16}O, the scale on the right to zero energy of the system $^{19}F+p$. Only transitions discussed in this report are shown. Bound levels of ^{20}Ne are omitted.

The angular momentum of the 482 kev. excited state in ^7Li has been the subject of considerable discussion (Hanna and Inglis 1949). It is formed in the radioactive decay of ^7Be, in the reaction $^{10}B(n, \alpha)^7Li$ and by inelastic scattering of protons and neutrons from lithium. Some of the evidence points to an angular momentum of $\frac{1}{2}$, while other data seem better fitted by $\frac{3}{2}$. It was pointed out (Feld 1949, Devons and Lindsey 1949) that the angular correlation between the alpha-particles and gamma-rays in the ^{10}B reaction would provide a sensitive test for these alternatives. Since the neutrons of thermal energy can only enter with $L = 0$, equations (8) and (9) are applicable and an isotropic correlation establishes $J' = \frac{1}{2}$ for the excited state. This is indeed the result found experimentally (Rose and Wilson 1950). The conclusion is confirmed by the isotropic distribution (Littauer 1950) of the gamma-rays from the inelastic

scattering of protons by lithium. Since the gamma-rays due to radiative capture by lithium at the same bombarding energy are not isotropic, as discussed above, the isotropy of the 482 kev. radiation is not due to s-wave protons or to a compound state with $J=0$ in ^{8}Be but to an excited state in ^{7}Li of $J'=\frac{1}{2}$.

(ii) *Radioactive Decay*

A considerable number of experiments have been performed on the angular correlations between radiations emitted in radioactive decays. The correlation between the alpha-particles and gamma-rays in the decay of RdTh shows a correlation involving terms in $\cos^2\theta$ and in $\cos^4\theta$ (Kulchitzki, Latychev and Buliginski 1949). Since both the initial and the final states are ground states of even–even nuclei they have probably zero angular momentum. The only possible assignment for the excited state is then $J'=2$, and the correlation should have the form $W(\theta)=\cos^2\theta-\cos^4\theta$. Because of background effects from other

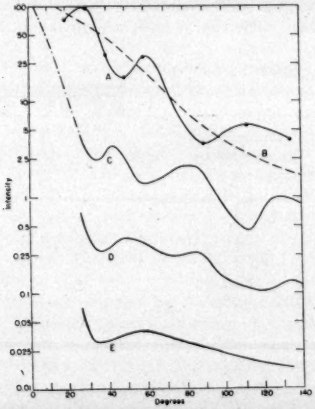


Figure 13. Angular distributions of long-range protons from ^{27}Al ^{28}(d, p) for deuteron energies of : A, 13·8 MeV.; C, 7·5 MeV.; D, 5·8 MeV.; E, 4·6 MeV. Curve A is from Gove (1950), C,D, E from Holt and Young (1950). B is a theoretical curve due to French and Goldberger, based on the 'stripping' model and calculated for 14 MeV. deuterons.

radiations it is not possible to make a quantitative comparison with the experimental results, but the expected maximum at 45° does appear in the data. If the assignment of $J'=2$ is correct, the excited state must have the same parity as the ground state and the radiation must be an electric quadrupole.

The suggestion that anisotropic correlations between successive gamma-rays might be expected was apparently first made by Dunworth (1940). Attempts to observe such correlations were unsuccessful until the use of the more efficient scintillation counters facilitated the experiments (Brady and Deutsch 1950). Positive results have been obtained in ten cases at the time of this writing. An interesting illustration is provided by the two gamma-rays following the beta-decay of the 5-yr. activity of ^{60}Co. The ground state of ^{60}Ni presumably has $J''=0$, which limits the number of theoretical possibilities (Hamilton 1940) with which the

data must be compared. The observed correlation is in good agreement with the form $W(\theta)=1+0{\cdot}125\cos^2\theta+0{\cdot}042\cos^4\theta$ (Brady and Deutsch 1950) predicted for $J=4$, $J'=2$, $J''=0$ if both gamma-rays are quadrupole quanta. By the method described in §3(iii), Figure 9, it has been possible to determine the correlation between the direction of polarization and angle of emission for these gamma-rays (Metzger and Deutsch 1950), and the result agrees with the above assignment, indicating further that both quadrupoles are electric. The conclusions from these experiments have found independent confirmation by the measurement (Waggoner, Moon and Roberts 1950) of the coefficients of internal conversion which indicate unambiguously that both transitions are electric quadrupoles. Together with this result the angular correlation and polarization experiments establish with certainty that the angular momenta of the two excited states in ^{60}Ni have angular momenta $J'=2$ and $J=4$ respectively and the same parity as the ground state. The fact that direct de-excitation, bypassing the intermediate state, occurs in less than $2{\cdot}5\times10^{-7}$ of the disintegrations (Fluharty and Deutsch 1949) is consistent with this.

Essentially identical results with those for ^{60}Co have been obtained by the same group of authors in the case of the decay of ^{46}Sc, leading to the same spin and parity assignments for the excited states of ^{46}Ti.

In other cases—^{134}Cs, ^{24}Na, ^{88}Y, ^{106}Rh (Brady and Deutsch 1950), ^{124}Sb and ^{110}Ag (Beyster and Wiedenbeck 1950)—no complete interpretation of the results has yet been possible. The correlation in the case of ^{88}Y shows that the first gamma-ray in the cascade must be a mixture of two L-values, probably electric quadrupole and magnetic dipole. This appears plausible in the light of the occurrence of direct transitions to the ground state in about two per cent of the transitions (Gamertsfelder 1944). The observed correlation in the case of ^{24}Na can be reconciled with the result of Rae (1949) that the 2·76 MeV. gamma-ray is an electric dipole only if the 1·38 MeV. transition is a mixture of L-values. There is, however, a serious difficulty in the fact that other evidence from Q-values of the inelastic scattering of protons and deuterons by magnesium shows a level at 1·38 MeV. but not at 2·76 MeV. Thus the 2·76 MeV. transition should precede the 1·38 MeV. gamma-ray. If the ground state of ^{24}Mg has angular momentum zero as expected, the 1·38 MeV. gamma-ray must then be a pure multipole. One link in this chain of paradoxical argument must be based on an erroneous experiment.

The angular correlation for ^{106}Rh seemed particularly interesting because allowed beta-transitions lead both to the ground state of ^{106}Pd and to the initial state of the two-quantum cascade. Since the ground state presumably has $J''=0$, a decision on the spin of the upper state could throw some light on the selection rules for allowed beta-decay. The observed correlation has the form $W(\theta)=1-1{\cdot}5\cos^2\theta+2\cos^4\theta$. Figures 14 and 15 show the angular and polarization correlations, respectively, for this case. No combination of two gamma-rays up to $L=3$, with or without interference terms, can explain this (Ling and Falkoff 1949, Spiers 1950). Spin assignments of $J=0$, $J'=2$, $J''=0$ yield a correlation of the observed form, but with coefficients twice as great as the observed ones. The polarization correlation (Metzger and Deutsch 1950) also has the form predicted by these spins assuming electric quadrupole radiations, but with the same discrepancy in the coefficients. The proposed assignment is probably the correct one. The decay scheme of ^{106}Rh is probably more complex than has been assumed, and other gamma-rays contribute an isotropic background to the correlation.

The decay of the 65-minute isomer of ^{204}Pb proceeds by successive emission of two gamma-rays. The intermediate state of this cascade has a lifetime of about 3×10^{-7} sec. Despite this long life, the correlation is not isotropic, showing that at least in this case atomic fields do not completely destroy the orientation of the nucleus in this time (Sunyar, Alburger, Friedlander, Goldhaber and Scharff-Goldhaber 1950). It has been proposed that an external magnetic field

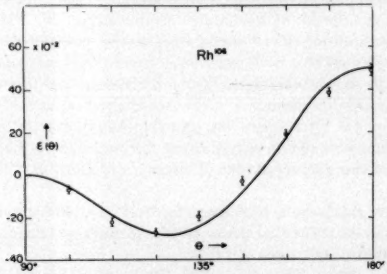


Figure 14. Angular correlation of gamma-rays from ^{106}Pd according to Brady and Deutsch (1950).

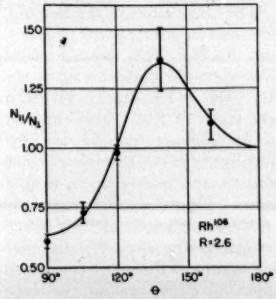


Figure 15. Polarization–direction correlation of gamma-rays from ^{106}P according to Metzger and Deutsch (1950).

perpendicular to the plane in which the correlation is measured will cause the angular pattern to precess with the nuclear Larmor frequency, permitting a measurement of the gyromagnetic ratio of the intermediate state (Sunyar, Alburger, Friedlander, Goldhaber and Scharff-Goldhaber 1950).

The theory of the angular correlation between beta-rays and gamma-rays is somewhat more complicated than the other cases discussed both because of the relativistic treatment of the electron and because an unobserved particle, the neutrino, participates in one of the transitions. The theory has been discussed by Falkoff and Uhlenbeck (1950 b). Equations (8) and (9) and the general conclusions from them are valid. Thus for allowed beta-transitions in which $L = 0$, according

to all plausible formulations of the theory all correlations must be isotropic. The same is true for some 'first forbidden' transitions. In general, anisotropy can be expected only if the shape of the beta-ray spectrum differs from the allowed Fermi distribution. The highest power of $\cos\theta$ which can appear is the same as the highest power of the electron momentum appearing in the 'correction factor' for the spectrum shape (Konopinski 1943). The magnitude of the anisotropy depends on the energy of the beta-ray, and is in general greatest for energies near the maximum. This is plausible if one considers the fundamental symmetry of electron and neutrino in the beta-decay process. The correlation between neutrino and gamma-ray must be very similar to that between electron and gamma-ray. In two cases the dependence of the correlation on beta-ray energy has been observed experimentally and the anisotropy found to increase with increasing energy. For ^{86}Rb the correlation for the most energetic electrons is approximately $W(\theta)=1+0{\cdot}2\cos^2\theta$ (Stevenson and Deutsch 1950), for ^{124}Sb approximately $W(\theta)=1-0{\cdot}5\cos^2\theta$ (Ridgway 1950, Beyster and Wiedenbeck 1950). In ^{170}Tm an anisotropic correlation has been observed (Novey 1950). The main interest of these experiments, besides the information concerning nuclear states involved, is the fact that the magnitude of the correlation and its energy dependence is determined by the nature of the interaction giving rise to the beta-decay process. The calculations (Falkoff and Uhlenbeck 1950 b) are incomplete since they do not take the effect of the nuclear Coulomb field into account, but progress has been made in this direction (Lennox 1950 and private communication). The experiments also need further refinement before definite conclusions can be drawn.

A more direct but technically much more difficult method to obtain information of this type is an observation of the angular correlation between electron and neutrino, or rather between electron and recoiling nucleus, in beta-decay. In this case anisotropy is expected even for allowed transitions because of the inherent admixture of p-wave components to the s-wave for a relativistic particle of half-integer spin. The correlation for allowed transitions is essentially an interference effect between s- and p-wave, and therefore contains only the first power of the cosine. Since the phenomenon is essentially relativistic, the coefficient of $\cos\theta$ is proportional to $v/\mathbf{c}$ for the electron. The magnitude and sign of the coefficient depend on the interaction giving rise to the transition. For the interaction originally proposed by Fermi, analogous to electromagnetic processes, the correlation between electron and neutrino should be essentially the same as that between electron and positron in internal pair conversion for 0–0 transitions as given in equation (15). Since the neutrino rest mass is zero, or very nearly so, the term in $m^2\mathbf{c}^4$ disappears. Also the ratio p/W is constant for the neutrino. Dropping constant terms, equation (15) becomes $W(\theta)=1+(p/W)\cos\theta$, where p and W refer to the electron. Thus for this particular interaction electron and neutrino tend to be emitted in the same direction, especially at high electron energies. This is not true for three of the other four possible interactions (Hamilton 1947). In forbidden transitions higher powers of $\cos\theta$ appear in the correlation.

Allen, Paneth and Morrish (1949) investigated the correlation in the case of the allowed beta-decay of ^{6}He. The beta-rays were observed in a Geiger-Müller counter and the recoil ^{6}Li ions, at a fixed angle with respect to the beta-rays, in an electron multiplier. The energy of the recoil ions was determined by a combination of retarding grids and a time-of-flight measurement. For a given electron energy

the recoil energy fixes the angle of the neutrino by virtue of the conservation of momentum and energy. The electron energy was not measured in Allen's experiment, so that the result represents an integral over the beta-ray spectrum. This fact reduces the sensitivity of the recoil energy distribution to the angular correlation. The result did indicate, however, fairly definitely that the tendency is for electron and neutrino to be emitted in *opposite* directions. Since the decay of ^{6}He almost certainly involves an angular momentum change of one, only the tensor or axial vector interactions can give rise to an allowed transition. Of these the axial vector theory predicts (Hamilton 1947) a correlation favouring opposite directions for electron and neutrino of the form $W(\theta) = 1 - \frac{1}{3}(p/W)\cos\theta$ while the tensor form requires a correlation favouring emission in the same direction. Allen's result seems to exclude the latter. Sherwin (1949) finds for ^{32}P that emission in the *same* direction is preferred. This is a forbidden transition and therefore much harder to interpret. Earlier results of the same author (1948) were shown by him to be in error. With some further improvement of technique angular correlation experiments of this type should definitely settle the question of predominant terms in beta-interaction.

REFERENCES

ALBURGER, D. E., and HAFNER, E. M., 1950, *Rev. Mod. Phys.*, **22**, 373.

ALLEN, J. S., PANETH, H. R., and MORRISH, A. H., 1949, *Phys. Rev.*, **75**, 570.

ARGO, H. V., GITTINGS, H. T., HEMMENDINGER, A., JARVIS, G. A., and TASCHEK, R. F., 1950, *Phys. Rev.*, **78**, 691.

ARNOLD, W. R., 1950, *Phys. Rev.*, **79**, 170.

AXEL, P., and DANCOFF, S. M., 1949, *Phys. Rev.*, **80**, 34.

BEIDUK, F. M., PRUETT, J. R., and KONOPINSKI, E. J., 1950, *Phys. Rev.*, **77**, 622, 628.

BELL, P. R., and JORDAN, W. H., 1949, *Nucleonics*, **5**, No. 4, 30.

DEBENEDETTI, S., COWAN, C. E., and KONNEKER, W. R., 1949, *Phys. Rev.*, **76**, 440.

BENNETT, W. E., BONNER, T. W., RICHARDS, H. T., and WATT, B. E., 1947, *Phys. Rev.*, **71**, 11.

BERINGER, R., 1943, *Phys. Rev.*, **63**, 23.

BERINGER, R., and MONTGOMERY, C. G., 1942, *Phys. Rev.*, **61**, 222.

BETHE, H. A., and BACHER, R. F., 1936, *Rev. Mod. Phys.*, **8**, 82.

BEYSTER, J. R., and WIEDENBECK, M. L., 1950, *Phys. Rev.*, **79**, 169, 411.

BLACKETT, P. M. S., and LEES, D. S., 1932, *Proc. Roy. Soc.* A, **136**, 338.

BONNER, T. W., and EVANS, J. E., 1948, *Phys. Rev.*, **73**, 666.

BOYER, K., 1950, *Ph.D. Thesis*, M.I.T.

BRADY, E. L., and DEUTSCH, M., 1950, *Phys. Rev.* **78**, 558.

CHADWICK, J., MAY, A. N., PICKAVANCE, T. G., and POWELL, C. F., 1944, *Proc. Roy. Soc.* A, **183**, 1.

CHAMPION, F. C., and ROY, R. R., 1948, *Phys. Rev.*, **74**, 5.

CHRISTY, R. F., COHEN, E. R., FOWLER, W. A., LAURITSEN, C. C., and LAURITSEN, T., 1947, *Phys. Rev.*, **72**, 698.

COCHRANE, W., and HESTER, A. G., 1949, *Proc. Roy. Soc.* A, **199**, 458.

COMPTON, A. H., and SIMON, A., W. 1925, *Phys. Rev.*, **26**, 289.

DANCOFF, S. M., and MORRISON, P., 1939, *Phys. Rev.*, **55**, 122.

DAY, R. B., CHAO, C. Y., FOWLER, W. A., and PERRY, J. E., 1950, *Bull. Amer. Phys. Soc.*, **25**, No. 4, 16.

DEVONS, S., 1949, *Proc. Phys. Soc.* A, **62**, 580.

DEVONS, S., HEREWARD, H. G., and LINDSEY, G. R., 1949, *Nature, Lond.*, **164**, 586.

DEVONS, S., and HINE, M. G. N., 1949, *Proc. Roy. Soc.* A, **199**, 56, 73.

DEVONS, S., and LINDSEY, G. R., 1949, *Nature, Lond.*, **164**, 539.

DRELL, S. D., 1949, *Phys. Rev.*, **75**, 132.

DUNWORTH, J. V., 1940, *Rev. Sci. Instrum.*, **11**, 167.

FALK, C. E., CREUTZ, E., and SEITZ, F., 1949, *Phys. Rev.*, **76**, 322.

FALKOFF, D. L., 1948, *Phys. Rev.*, **73**, 518.
FALKOFF, D. L., and UHLENBECK, G. E., 1950 a, *Phys. Rev.*, **79**, 323; 1950 b, *Ibid.*, **79**, 334.
FEATHER, N., and DUNWORTH, J. V., 1938, *Proc. Roy. Soc.* A, **168**, 566.
FELD, B. T., 1949, *Phys. Rev.*, **75**, 1618.
FLUHARTY, R. G., and DEUTSCH, M., 1949, *Phys. Rev.*, **76**, 182.
FOWLER, W. A., LAURITSEN, C. C., and TOLLESTRUP, A. V., 1949, *Phys. Rev.*, **76**, 1767.
GAMERTSFELDER, G. R., 1944, *Phys. Rev.*, **66**, 288.
GENEVESE, F., 1949, *Phys. Rev.*, **76**, 1288.
GILBERT, C. W., 1948, *Camb. Phil. Soc.*, **44**, 447.
GOERTZEL, G., 1946, *Phys. Rev.*, **70**, 897.
GOOD, W. M., 1946, *Phys. Rev.*, **70**, 978.
GORDY, W., 1948, *Rev. Mod. Phys.*, **20**, 668.
GOVE, H. E., 1950, *Ph.D. Thesis*, M.I.T., also *Phys. Rev.*, **78**, 345.
GRAHAM, G., and HALBAN, H., 1945, *Rev. Mod. Phys.*, **17**, 297.
HAMILTON, D. R., 1940, *Phys. Rev.* **58**, 122; 1947, *Ibid.*, **71**, 456; 1948, *Ibid.*, **74**, 782.
HANNA, S. S., and INGLIS, D. R., 1949, *Phys. Rev.*, **75**, 1767.
HEBB, M. H., and NELSON, E., 1940, *Phys. Rev.*, **58**, 486.
HEYDENBURG, N. P., and INGLIS, D. R., 1948, *Phys. Rev.*, **73**, 230.
HEYDENBURG, N. P., INGLIS, D. R., WHITEHEAD, W. D., Jr., and HAFNER, E. M., 1949, *Phys. Rev.*, **75**, 1147.
HOLT, J. R., and YOUNG, C. T., 1950, *Proc. Phys. Soc.* A, **63**, 833.
HORNYAK, W. F., LAURITSEN, T., MORRISON, P., and FOWLER, W. A., 1950, *Rev. Mod. Phys.*, **22**, 291.
INGLIS, D. R., 1948, *Phys. Rev.*, **74**, 21.
JACOBS, J. A., MALMBERG, P. R., and WAHL, J. S., 1948, *Phys. Rev.*, **73**, 1130.
JACOBS, J. A., and WILSON, W. L., 1941, *Phys. Rev.*, **59**, 108.
KELLOGG, J. B. M., and MILLMAN, S., 1946, *Rev. Mod. Phys.*, **78**, 323.
KIKUCHI, S., WATASE, Y., and ITOH, J., 1942, *Z. Phys.*, **119**, 185.
KOJIMA, S., 1943, *Proc. Imp. Acad. Tokyo*, **19**, 282.
KONOPINSKI, E. J., 1943, *Rev. Mod. Phys.*, **15**, 209.
KULCHITZKI, L. A., LATYCHEV, G. D., and BULIGINSKI, D. G., 1949, *Acad. Nauk. U.S.S.R.* (*Doklady*), **64**, 57.
LASSEN, N. O., 1949, *Phys. Rev.*, **75**, 1099.
LENNOX, E. S., 1950, *Phys. Rev.*, **79**, 221.
LING, D. S., and FALKOFF, D. L., 1948, *Phys. Rev.*, **74**, 1224; 1949, *Ibid.*, **76**, 431.
LITTAUER, R. M., 1950, *Proc. Phys. Soc.* A, **63**, 294.
LIVESEY, D. L., and WILKINSON, D. H., 1948, *Proc. Roy. Soc.* A, **195**, 122.
MAIER-LEIBNITZ, H., 1942, *Z. Phys.*, **43**, 333; 1946, *Z. Naturforsch.*, **1**, 243.
MAYER, M. G., 1950, *Phys. Rev.*, **78**, 16, 22.
MEINERS, E. P., 1949, *Phys. Rev.*, **76**, 259.
METZGER, F., and DEUTSCH, M., 1950, *Phys. Rev.*, **78**, 537.
MITCHELL, A. C. G., 1950, *Rev. Mod. Phys.*, **22**, 36.
NEUERT, H., 1937, *Phys. Z.*, **38**, 618.
NOVEY, T. B., 1950, *Phys. Rev.*, **78**, 66.
OPPENHEIMER, J. R., 1941, *Phys. Rev.*, **60**, 164.
OPPENHEIMER, J. R., and SCHWINGER, J., 1939, *Phys. Rev.*, **56**, 1066.
PERLMAN, I., GHIORSO, A., and SEABORG, G. T., 1950, *Phys. Rev.*, **77**, 26.
POPPEMA, O. J., 1950, *Helv. Phys. Acta*, **23**, Supp. III, 189.
POWELL, C. F., and OCCHIALINI, G. P. S., 1947, *Nuclear Physics in Photographs* (Oxford : Clarendon Press).
RAE, E. R., 1949, *Phil. Mag.*, **40**, 1155.
RARITA, W., and SCHWINGER, J., 1941, *Phys. Rev.*, **59**, 436.
RASMUSSEN, V. K., HORNYAK, W. F., LAURITSEN, C. C., and LAURITSEN, T., 1950, *Phys. Rev.*, **77**, 617.
RESNICK, I., and HANNA, S. S., 1949, *Phys. Rev.*, **76**, 168.
RESNICK, R., and INGLIS, D. R., 1949, *Phys. Rev.*, **76**, 1318.
RIDGWAY, S. L., 1950, *Phys. Rev.*, **78**, 821.
ROSE, B., and WILSON, A. R. W., 1950, *Phys. Rev.*, **78**, 68.
ROSE, H. E., 1949, *Phys. Rev.*, **76**, 678.

ROSE, M. E., GOERTZEL, G. H., SPINRAD, B. I., HARR, J., and STRONG, P., 1949, *Phys. Rev.*, **76**, 1883; also "Table of K Shell Internal Conversion Coefficients" (report privately circulated).
SEGRÈ, E., and HELMHOLZ, A. C., 1949, *Rev. Mod. Phys.*, **21**, 271.
SERBER, R., 1947, *Phys. Rev.*, **73**, 1007.
SHERWIN, C. W., 1948, *Phys. Rev.*, **73**, 216, 1173; 1949, *Ibid.*, **75**, 1799.
SHULL, F. B., and FEENBERG, E., 1949, *Phys. Rev.*, **75**, 1768.
SPIERS, J. A., 1948, *Nature, Lond.*, **161**, 807; 1950, *Phys. Rev.*, **78**, 75.
STEVENSON, D. T., and DEUTSCH, M., 1950, *Phys. Rev.*, **78**, 640.
SUNYAR, A. W., ALBURGER, D., FRIEDLANDER, G., GOLDHABER, M., and SCHARFF-GOLDHABER, G., 1950, *Phys. Rev.*, **79**, 181.
TASCHEK, R. F., 1948, *Rev. Sci. Instrum.*, **19**, 591; 1949, *Phys. Rev.*, **76**, 584.
VAN ALLEN, J. A., and SMITH, N. M., 1941, *Phys. Rev.*, **59**, 501.
WAGGONER, M. A., MOON, M. L. and ROBERTS, A., 1950, *Phys. Rev.*, **80**, 420
WHALING, W., and BONNER, T. W., 1950, *Phys. Rev.*, **79**, 258.
WHALING, W., EVANS, J. E., and BONNER, T. W., 1949, *Phys. Rev.*, **75**, 688.
WILKINS, T. R., 1941, *Phys. Rev.*, **60**, 365.
WOODWARD, W. M., and HALPERN, I., 1949, *Phys. Rev.*, **76**, 107.
YAGODA, H. J., 1949, *Radioactive Measurements with Nuclear Emulsions* (New York: Wiley).
YANG, C. N., 1948, *Phys. Rev.*, **74**, 764.

THE NUCLEAR INTERACTIONS OF COSMIC RAYS

By G. D. ROCHESTER and W. G. V. ROSSER
The Physical Laboratories, University of Manchester

CONTENTS

PAGE

I. Introduction 228
II. Types of nuclear interaction 229
§ 1. Evaporation stars 229
1.1. Star types 229
1.2. General features of evaporation stars 231
1.3. Nuclear evaporation 235
1.4. Evaporation stars in electron-sensitive emulsions 238
1.5. Heavy fragments 240
§ 2. Penetrating showers 241
2.1. The investigation of penetrating showers by counter arrangements and the cloud chamber 241
2.2. Penetrating showers in electron-sensitive emulsions 242
2.3. The determination of the momentum spectrum of shower particles by the cloud chamber 248
2.4. Mixed showers 249
2.5. Heavy unstable particles in penetrating showers 252
§ 3. The nuclear interactions of heavy nuclei 253
§ 4. The nuclear interactions of the non-nucleonic components of cosmic rays 254
4.1. Electrons and photons 254
4.2. μ-mesons 254
4.3. π-mesons 256
III. The transition effects of the nucleonic and nuclear components 257
§ 1. Interaction mean free path and attenuation lengths 257
§ 2. The interaction mean free paths of the heavy primaries 259
§ 3. The interaction mean free paths of the primaries and secondaries of penetrating showers 259
3.1. Introduction 259
3.2. The transition effect of penetrating showers 260
3.3. The determination of the interaction mean free path of charged primaries 260
3.4. The determination of the interaction mean free path of neutral primaries 262
3.5. The secondary interactions of the particles produced in stars in the photographic emulsion 264
3.6. The large-angle scatterings and the secondary interactions of the penetrating particles observed in the cloud chamber 265
§ 4. The attenuation of high-energy nucleons in the atmosphere 265
4.1. Introduction 265
4.2. Counter experiments on the variation of penetrating showers with altitude 267
4.3. Photographic-plate experiments on the variation of high-energy stars with altitude 267
4.4. The barometer effect of penetrating showers 268
4.5. The angular distribution of the primaries of penetrating showers 269

§ 5. The attenuation of star-producing nucleons in the atmosphere and in condensed matter 269
5.1. Introduction 269
5.2. Photographic-plate experiments on the variation of evaporation stars with altitude 269
5.3. Ionization-chamber experiments on the variation of low-energy stars with altitude 270
5.4. Transition effects in condensed matter 270
§ 6. The latitude effects of penetrating showers and stars 271
6.1. Penetrating showers 271
6.2. Stars 271
IV. The nucleonic components in the atmosphere and their correlation 272
§ 1. Introduction 272
§ 2. The primaries of penetrating showers : the A component 273
2.1. Exponential absorption 273
2.2. Meson production 273
§ 3. Single protons 275
§ 4. The primaries of evaporation stars : the B component 277
§ 5. The correlation of different types of stars 278
§ 6. Slow protons and slow neutrons : the C component 279
6.1. Slow protons 279
6.2. Slow neutrons 280
V. Conclusion 283
Acknowledgments 285
References 285

ABSTRACT. This report deals with the types of nuclear interaction, the properties of nuclear-interacting particles, and the correlation of nuclear events in the atmosphere and in condensed matter. The empirical data for evaporation stars are collected and it is shown that the statistical theory of the nucleus can account for much of the data. The methods of investigating meson showers are outlined and the principal results given. Recent work on the new heavy unstable particles is included in this section. An account is then given of the nuclear interaction of non-nucleonic particles, i.e. electrons, photons, μ-mesons and π-mesons.

In the next section experiments which give information about the interactions of nucleons in the atmosphere and in condensed matter are discussed. The meaning of interaction mean free path and attenuation length is brought out. Finally, the nucleonic components in the atmosphere and their correlation are discussed in detail. It is shown that the formation of stars of various types in the atmosphere, and the equilibrium of the various components in the lower reaches of the atmosphere, can be understood on the picture of nucleon cascades which start at the top of the atmosphere and end with the formation of slow neutrons and protons.

The report concludes with a discussion of some of the main problems to be solved in this field of research.

I. INTRODUCTION

COSMIC rays at sea level show two distinct components: the so-called soft and hard components. The soft component is now known to consist of electrons and photons, and the hard component of μ-mesons. While these components constitute the main part of cosmic radiation at sea level, there is clear evidence for the presence of other types of particle. Both the cloud chamber and the photographic emulsion show the presence of single identifiable protons, and neutron counters reveal the presence of slow neutrons.

Electrons and photons interact with matter through long-range forces; neutrons and protons, however, interact through short-range forces whose

nature is still only partially understood. It is the purpose of this article to review the nuclear interactions which these nucleonic radiations undergo in the earth's atmosphere and in condensed matter. These interactions lead to the formation of nuclear disintegrations or stars, so that a review of the nuclear-interacting radiation in cosmic rays is largely a review of the experimental evidence relating to stars. It will appear that the particles constituting the main part of the nuclear interacting radiation are protons and neutrons, but evidence will be presented for the nuclear interaction of other types of particle.

In Part II of this article the main types of nuclear interaction will be reviewed. This part will be mainly concerned with low-energy interactions, or evaporation stars, in which the excitation energy of the nucleus is of the order of a few hundred million electron volts (MEV.), and high-energy interactions, or penetrating showers, which are nuclear events in which several mesons are created. Since the mass of the created meson (which is believed to be the π-meson) is 280 m_e, where m_e is the mass of the electron, a minimum energy of several hundred MEV. is necessary for the production of a penetrating shower. Also discussed in this part will be the nuclear interaction of the heavy nuclei which are observed near the top of the atmosphere.

In Part III of the article the transition effects of the nucleonic and nuclear components in the atmosphere and in condensed matter will be considered. The discussion will lead in Part IV to an account of how the nuclear interacting component of cosmic radiation builds up by nucleon cascades in the atmosphere and of how mesons are created.

Two review articles by Rossi (1948, 1949), which cover most of the work published up to January 1949 on the subject of the nuclear interacting rays of cosmic radiation, have already appeared. In the present article an attempt is made to include work published up to June 1950. C. F. Powell has given a review of recent work on mesons in Volume XIII of these Reports.

II. TYPES OF NUCLEAR INTERACTION

§ 1. EVAPORATION STARS

1.1. *Star Types*

The first extensive series of experiments on nuclear disintegrations induced by cosmic rays was made by Blau and Wambacher in 1937. These workers found that if unexposed Ilford half-tone plates were left for several months at mountain altitudes they showed on development star-like events which were clearly the disintegration products of nuclear explosions. The rapid increase of such stars with altitude (Stetter and Wambacher 1939) proved that they were connected with the cosmic radiation, and the rate of rise was so similar to the increase in the soft component that it was assumed that the stars were connected with this component. The rapid increase with altitude definitely excluded the hard components as a possible source of the stars.

The application of the cloud chamber to this problem, by Hazen (1944) and W. M. Powell (1946), made it certain that the soft component was not the main source of the stars. Hazen operated a cloud chamber containing several lead plates at an altitude of 3,000 m. and in 8,500 photographs taken at random found 58 examples of stars. Powell in 20,000 photographs taken at 4,300 m. found 156 stars in the lead plates and 13 stars in the gas. The main results of

these researches showed that the stars were of two kinds: (i) low-energy stars whose tracks consisted mainly of protons of a few million electron volts energy, and (ii) high-energy stars whose tracks consisted of a mixture of low and high energy protons, α-particles, and other types of particle. Stars of type (i) seemed to be produced entirely by a neutral radiation which was different from γ-radiation since it was not accompanied by electron–photon cascade showers, and stars of type (ii) equally by charged and uncharged primaries. Stars of type (i) were much more frequent than those of type (ii). Since the only nuclear-interacting neutral particle known was the neutron it was assumed that the low-energy stars were caused by cosmic-ray neutrons. The high-energy stars were assumed to be due to fast protons and neutrons. Later work has upheld this conclusion, although it is now known that a small percentage of stars is due to other types of particle.

Heisenberg in 1937 suggested that stars of type (i) would contain two main types of particle, namely, fast knock-on nucleons and Bohr 'evaporation' particles. In the Heisenberg model the primary particle and its fast secondary products were envisaged as causing a 'heating up' of the nucleus by collision with other nucleons within the nucleus, the excited nucleus subsequently returning to a stable state by a process analogous to evaporation. The statistical treatment of nuclear disintegration was initiated by Bohr and Kalckar (1937) and amplified by Bethe (1937) and Weisskopf (1937). It was applied in detail by Bagge (1941, 1943) to the results of Wambacher and her collaborators and shown to account for the main features of their stars. It was concluded that the main process by which stars were formed was the evaporation of nuclear particles from an excited nucleus. Bagge assumed that the soft component was responsible for this 'excitation', but it is now known from the cloud-chamber work already referred to, and from further work with photographic plates, that neutrons are the primary agent. Protons are unlikely to produce these stars because at these energies the loss of energy by ionization is so great that very few exist in the cosmic radiation compared with the numbers of neutrons. Such stars are now commonly called 'evaporation stars'. The kinetic energies of the protons produced in evaporation seldom exceed 30 Mev. An example of a large evaporation star is shown in Plate I*. Heisenberg knock-on nucleons may also be present in these stars, but they may or may not be observed, depending on the sensitivity of the recording device. A cloud chamber will usually record all types of particle, but a photographic emulsion may be chosen so as to record only one type of particle. Thus, although a fully developed Ilford C2 emulsion will record protons of energies up to 100 Mev. it is possible to develop this emulsion to be sensitive only to protons of energy less than 40 Mev.† (Page 1950). This emulsion would then record mainly 'evaporation protons', and not protons of higher energy. It would, of course, also record α-particles, deuterons etc. of certain energies, but again these would be of the 'evaporation' type.

There remain the high-energy stars. These are complex stars consisting of evaporation particles, knock-on nucleons and mesons. Clearly the main difference between this type of star and the one already described is the additional process of meson formation. Since we are now in an energy region ($>$1,000 Mev.) in which the proton loses only a small fraction of its energy by ionization, this type of star can be produced equally by neutrons and protons.

* Plates are printed together opposite p. 240.

† In this energy region the ionization produced by a proton decreases with increasing energy.

1.2. *General Features of Evaporation Stars*

We proceed now to consider some further features of evaporation stars which have resulted from work carried out with nuclear research emulsions. These features include the size–frequency distribution of stars, the types of particle emitted and their energy distributions, and the angular distributions of the various kinds of particle. The transition effects of stars will be considered in Part III.

(a) The size–frequency distribution of stars.

The integral size distribution of stars has been found by many workers, and the results of Page (1950) using Ilford C2 nuclear research emulsion at an altitude of 3,450 m. will be taken as typical. A plot of the results on a semi-logarithmic

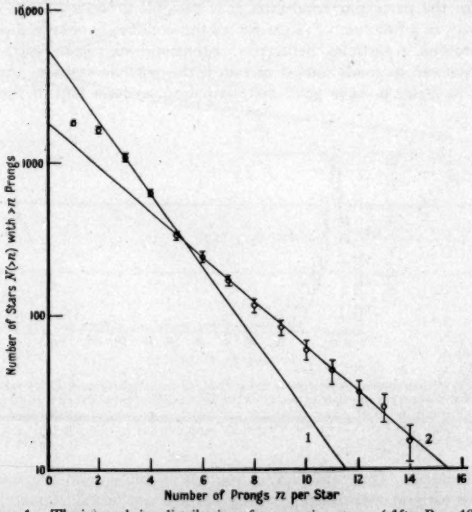


Figure 1. The integral size distribution of evaporation stars. (*After* Page 1950)

scale (Figure 1) shows two straight lines which are given by the expressions

$$\left.\begin{aligned} N(>n) &= A\exp(-0{\cdot}521n) \quad \text{for} \quad n \leqslant 6, \\ \text{and} \qquad N(>n) &= 0{\cdot}38A\exp(-0{\cdot}331n) \quad \text{for} \quad n>6. \end{aligned}\right\} \qquad \ldots\ldots(1.21)$$

The decrease in the rate of stars with number of prongs is an indication of the preponderance of low-energy star primaries in the atmosphere. The change in slope of the curve at $n=6$ may be due to the fact that the light elements present in the emulsion contribute only to the small stars. Barton, George and Jason (1951) have recently shown that the plot of the frequency of occurrence of stars (i.e. $N(E)\,dE$) against the logarithm of the total energy released in the star gives a good straight line throughout the whole range of star sizes. The results fit a power law of the form

$$N(E)\,dE = KdE/E^{\gamma} \qquad \text{with} \qquad \gamma = 2{\cdot}65 \pm 0{\cdot}1.$$

According to Bernardini, Cortini and Manfredini (1950), who obtain, at the same altitude, results closely similar to Page, the average number of evaporation prongs per star of three or more prongs is 4·7.

(*b*) *The types of particle emitted from stars and their energy distributions.*

The particles emitted from a star are identified by the examination of the tracks which they leave in the emulsion. The two quantities which usually give most information are the grain density and the range, but in certain cases, particularly for mesons, the scattering of the particle can also be used. By the use of calibration particles, for example, protons and α-particles of known energy which may be put into the emulsion artificially, and the known range–energy relations for the particular emulsions, it is possible to draw a series of curves of grain density as a function of range for all the commonly observed particles (for example protons, α-particles, deuterons, tritons and the various sorts of mesons). Identification can be made almost certain if the particle stops in the emulsion; otherwise, in order to have good discrimination, at least half of the true range

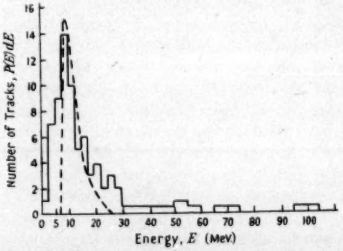


Figure 2. Energy distribution of 80 proton tracks from 15 disintegrations of heavy nuclei (Ag or Br). Ordinate denotes the number of observed tracks per 2·5 MeV. for the energy interval 0–30 MeV. and half the number of observed tracks per 5 MeV. for energies greater than 30 MeV. The broken curve corresponds to the Weisskopf equation (of § 1.3 in Part II) with $T=2{\cdot}5$ MeV. and $V=7$ MeV. (*After* Perkins 1947 a)

must be observed. Once the type of particle is known and the residual range measured or estimated the energy of the particle can be calculated.

Great care is necessary in the choice of a suitable emulsion and in the development of the emulsion. The recording characteristics of the emulsion represent a compromise between sensitivity on the one hand and discrimination on the other. The reason why high sensitivity and high discrimination do not occur together is that the former demands the highest possible grain density for a given specific ionization while the latter requires a low grain density. As the sensitivity is increased a point is reached at which practically no discrimination between particles of the same charge is possible because the developed grains fuse together and form a continuous line.

As will be clear from §1.1 the Ilford C2 nuclear research emulsion is very suitable for the study of evaporation particles. This emulsion and other similar emulsions have been applied with success to this problem by several groups of workers, notably Perkins (1947 a), Harding, Lattimore and Perkins (1949) Page (1950) and Bernardini, Cortini and Manfredini (1950). Although the methods of

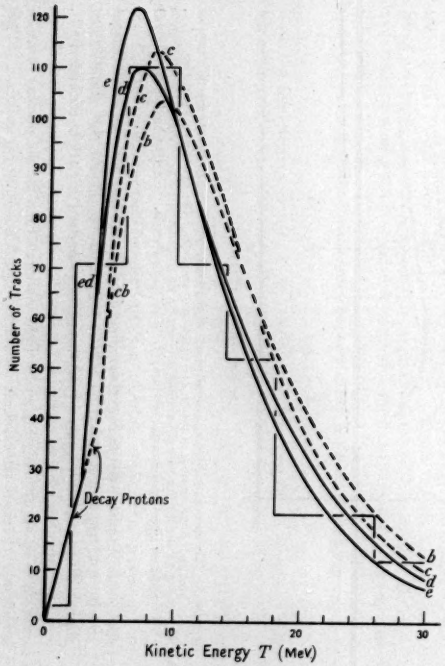


Figure 3. Theoretical and experimental results for the energy spectrum of particles of charge e evaporated from stars with an average of 10·2 prongs. The four theoretical curves b, c, d and e are for four different sets of parameters. The parameters for d and e are given in equations (1.34) and (1.35) of § 1.3 in Part II. (*After* Le Couteur 1950 a)

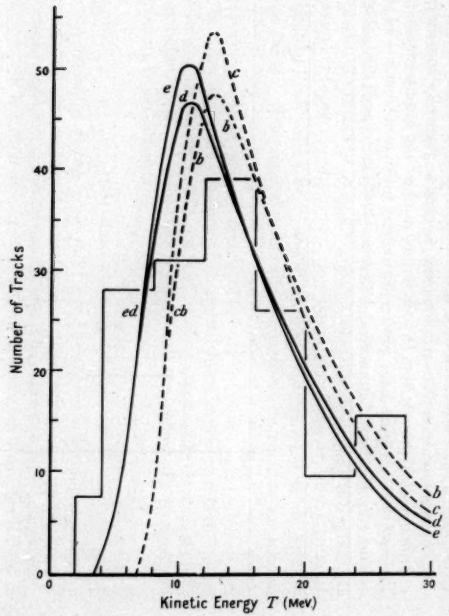


Figure 4. Theoretical and experimental results for the energy spectrum of evaporated particles of charge $2e$ from stars with an average of 10·2 prongs. The four theoretical curves b, c, d and e are for four different sets of parameters. The parameters for d and e are given in equations (1.34) and (1.35) of § 1.3 in Part II. (*After* Le Couteur 1950 a)

identification have varied slightly (for example, the Italian workers require that all particles to be identified shall end in the emulsion), the results are very similar. Typical sets of results are shown in Figures 2, 3, 4 and 5. Figure 2 shows the energy distribution of 80 protons from 15 stars formed in silver and bromine measured by Perkins (1947 a), Figure 3 the proton distribution for stars of from 6 to 14 prongs as found by Harding, Lattimore and Perkins (1949) and Page (1950), Figure 4 the energy distribution of α-particles for stars of from 6 to 14 prongs as measured by Harding, Lattimore and Perkins, and Figure 5 the results of Page on the ratio of the number of tracks of particles with charge $2e$ to all particles as a function of

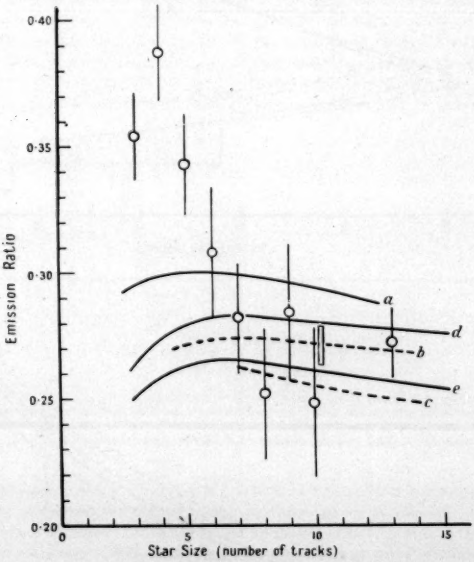


Figure 5. Theoretical and experimental values of the emission ratio, i.e. the number of tracks of charge $2e$/total number of tracks, as a function of star size. The five theoretical curves *a*, *b*, *c*, *d* and *e* are for five sets of parameters given by Le Couteur (1950 a). The parameters for *d* and *e* are given in equations (1.34) and (1.35) of § 1.3 in Part II. (*After* Le Couteur 1950 a)

star size. Similar results have been found by Bernardini and his collaborators and by Addario and Tamburino (1949). Perkins' results for this quantity are slightly different, but they are within three times standard deviation (Perkins 1950).

The ratio of number of particles of charge $2e$ to number of particles with charge e seems to be appreciably higher for light materials than for the nuclear research emulsion. Perkins (1949) finds a value approaching unity for gelatine. This probably indicates that light elements like carbon and oxygen tend to break up into α-particles.

As regards particles other than protons and α-particles produced in the emulsion, Harding, Lattimore and Perkins find that approximately 30% of all singly charged particles are heavier than protons. These particles are mainly deuterons and tritons.

(*c*) *The angular distribution of evaporation particles.*

It is to be expected that in a pure evaporation process the particles will be emitted isotropically in the centre of mass system. Harding, Lattimore and Perkins (1949) have examined this point for 78 stars which must have arisen in either silver or bromine nuclei. They find a spherically symmetrical distribution for protons from stars of sizes 7–22 prongs and for α-particles from stars of sizes 7–14 prongs. A slight departure from spherical symmetry is, however, observed in the α-particles from the larger stars. It is suggested, somewhat speculatively, that local heating of the nucleus might lead to the ejection of some α-particles near the surface of the nucleus before thermal equilibrium is established throughout the nucleus and general evaporation sets in.

1.3. *Nuclear Evaporation*

According to the picture given in § 1.1 the passage of a fast nucleon across a heavy nucleus leads to a general heating up, or state of excitation, of the nucleus. If the excitation energy U of this excited nucleus is less than its total binding energy, the nucleus will not explode. Instead, it will behave somewhat like a fluid in thermal equilibrium. The nucleons of which the nucleus is composed will collide with one another much as do the molecules in a gas or a liquid. Again, in much the same way as in the evaporation of a molecule from a liquid, one of the nucleons may acquire, by statistical fluctuation, more energy than the average, and be able to leave the nucleus. This process will go on until the nucleus reaches a stable state. Since the energy levels of a heavy, highly excited nucleus are close together, thermodynamical reasoning can be applied to determine the distribution of energy among the nuclear particles. Weisskopf (1937) showed that in this case the probability of a neutron being emitted with energy between E and $E+dE$ is of the form

$$P(E)\,dE=(E/T^2)\exp(-E/T)\,dE. \qquad \ldots\ldots(1.31)$$

Both the nuclear temperature T and the energy E are measured in MeV. The temperature T is related to the excitation energy U by the equation

$$U=\beta T^n, \qquad \ldots\ldots(1.32)$$

where β is a constant involving the atomic weight of the nucleus and n depends on the assumed interaction between the nuclear particles (for example liquid-drop model, gas model, etc). In the liquid-drop model, for fairly high values of U (i.e. U of the order of several hundred MeV.), $n\simeq 4$, while in the Fermi-gas model, in which the nucleons are assumed to be an assemblage of free particles (with kinetic energy very much greater than the interaction energy), $n=2$.

In the case of charged particles equation (1.31) becomes

$$P(E)\,dE=[(E-V)/T^2]\exp[-(E-V)/T]\,dE \qquad \ldots\ldots(1.33)$$

for $E>V$. V is the height of the potential barrier, assumed impenetrable. The effect of the barrier is thus simply to shift the whole energy distribution by an amount V. The average energy of the emitted particles is $\bar{E}=2T+V$. The distribution

given by equation (1.33) for $T=2{\cdot}5$ Mev. and $V=7$ Mev. is shown by the broken line in Figure 2, where it is compared with the experimental results of Perkins (1947 a). It can be seen that while the general form of the curve is approximately correct there are many protons of energies greater than 30 Mev. and less than 7 Mev. which are not accounted for. The high-energy particles are undoubtedly the protons resulting from the Heisenberg knock-on process, and clearly they cannot be accounted for on an evaporation theory. In the bigger stars approximately 5–10% of all the particles observed in Ilford C2 plates are of this type. The low-energy protons will be considered later.

Equation (1.33) is true for the energy distribution of molecules evaporating from a drop of water, but it is not strictly true for the disintegration of a nucleus because the number of particles in the nucleus ($\sim$100) is so small that it cools appreciably after the emission of each particle. The effect of cooling has been allowed for by Harding, Lattimore and Perkins (1949) and they have shown that, especially with the Fermi-gas model, an appreciable improvement in the fit with the experimental curve at the low-energy side of the distribution is obtained. Even so there is still an excess of low-energy protons; the same feature is found in the α-particle distribution. Harding, Lattimore and Perkins suggest that these additional low-energy particles are due mainly to the quantum mechanical 'tunnel' effect or barrier penetration first suggested by Gamow to explain the emission of low-energy α-particles from radioactive atoms. This effect is undoubtedly present (especially for α-particles), but another process suggested independently by Fujimoto and Yamaguchi (1949) and Le Couteur (1950 a) is probably more important for protons. These authors suggest that in the later stages of cooling the nucleus emits many more neutrons than protons, thus increasing the ratio of the atomic charge to the atomic number. The nucleus then becomes unstable to the loss of positive charge, and the emission of 'decay protons' becomes more probable than the emission of positive beta-particles. The lifetime of this process is found to be between 10^{-15} and 10^{-3} sec. for a nucleus of atomic weight 100, and for the emission of protons in the energy range $1{\cdot}5-3{\cdot}0$ Mev. This process therefore increases the number of low-energy protons.

The work of Harding, Lattimore and Perkins has recently been extended by Le Couteur (1950 a) in a comprehensive theory of nuclear evaporation. Le Couteur has used the Fermi-gas model and has applied the theory only to a carefully selected range of star sizes. The lower limit of size is taken as 7 prongs in order to avoid complications from the light elements present in the photographic emulsion, and the upper limit is taken as 14 prongs in order that the excitation energy shall be less than the total binding energy. The calculations are made for excitation energies in the range $100-600$ Mev. Besides the effects allowed for by Harding, Lattimore and Perkins, Le Couteur has taken into account the thermal expansion of the nucleus and the neutron excess. He has obtained the energy distributions of the evaporated particles and the expected numbers of evaporated neutrons, protons, deuterons, ^{3}H, ^{3}He and α-particles as a function of the excitation energy of the nucleus. The results are compared with the combined experimental results of Harding, Lattimore and Perkins, and Page in Figures 3, 4, and 5. The various curves represent slightly different sets of parameters. Typical parameters are those for the curves marked *d* and *e*, for which

$$\beta=\tfrac{1}{4}A/3{\cdot}3 \ (d) \quad \text{and} \quad \beta=\tfrac{1}{4}A/2{\cdot}9 \ (e) \qquad \ldots\ldots\ldots(1.34)$$

For both sets of parameters the potential barriers of protons (p), deuterons (d), tritons (t) and α-particles are given by

$$V_p' = 0{\cdot}7V;\ V_d' = 0{\cdot}77V;\ V_t' = 0{\cdot}8V;\ V_\alpha' = 1{\cdot}6V \quad \text{and} \quad V = \frac{6}{1+U/200}\ \text{Mev.} \qquad \ldots\ldots(1.35)$$

In the last expression of (1.35) Le Couteur assumes with Bagge that the potential barrier is reduced at high excitation energies. It does not seem possible to account for all the low-energy α-particles unless an assumption of this kind is made. Figure 6 shows Le Couteur's results for the relative numbers of the different types of particle as a function of excitation energy.

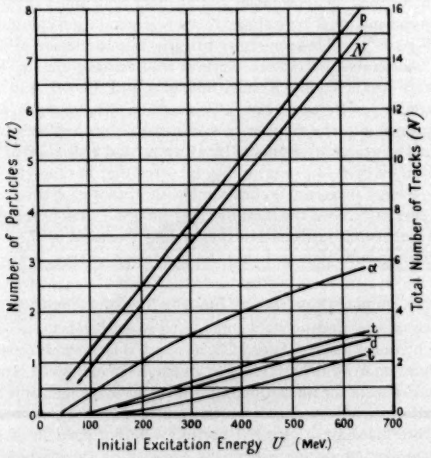


Figure 6. The total number N of tracks with numbers n of protons, deuterons, tritons and α-particles evaporated from a nucleus with initial excitation energy U. Parameters: mean of d and e. (*After* Le Couteur 1950 a)

Another test of the Le Couteur theory is the relationship between the mean energy of the emitted particles and the star size. According to the theory (Le Couteur 1950 b) the mean energy should increase almost linearly with star size. The results with parameters d and e give good agreement with the experimental results of Harding, Lattimore and Perkins, and Page for protons. There are not sufficient experimental data to check the calculations for α-particles.

Perkins (1950) has observed that the minimum energy of the α-particles emitted from stars decreases from 10 Mev. for small stars of two or three prongs to 2 Mev. for stars with more than 12 prongs. This difference is ascribed by Le Couteur to the reduction in the potential barrier predicted by equation (1.35). The corresponding effect for protons is completely masked by the 'decay protons' which set the lower limit to the energy spectrum.

So far the discussion has been restricted to stars of from 7 to 14 prongs in size, and it is now necessary to discuss stars outside this range. Stars of size less than 7 prongs include small stars produced in heavy nuclei, for example in silver and bromine, and stars produced in the light elements carbon, nitrogen and oxygen. Harding (1949 b) and Perkins (1950) have investigated the contribution from the light elements by studying the formation of stars in emulsions made up of alternate layers of nuclear research emulsion and gelatine. In this way Harding found that $36 \pm 4{\cdot}5\%$ of all stars are formed in light nuclei, a figure which agrees approximately with that calculated on the assumption that relative numbers of stars formed in the different elements in the emulsion are proportional to their geometrical cross sections. Thus it is fair to assume that some 35% of all stars with less than 7 prongs will be formed in light nuclei. The other 65% must then be formed in silver and bromine and must correspond to average excitation energies of less than 300 Mev. (see Figure 6). These energies bring us into the region of stars produced by artificially accelerated particles. Several studies have already been made at Berkeley of stars produced by 90 Mev. neutrons and 35, 90, 130 and 190 Mev. deuterons (Gardner and Peterson 1949). With deuterons, for example, the average star size is 3 prongs and the frequency–size distribution has been accounted for qualitatively on the evaporation model by Horning and Baumhoff (1949).

For stars of more than 14 prongs it may be noted that Perkins (1950) has shown that the proton energy spectrum is consistent with evaporation theory even though the excitation energy exceeds the total binding energy. This probably means that before most of the particles are emitted the cooling produced by the emission of the first few particles is sufficient to bring the excitation energy below the binding energy.

1.4. *Evaporation Stars in Electron-Sensitive Emulsions*

All the work on evaporation stars considered so far has been carried out using nuclear research emulsions, e.g. Ilford C2, in which it is possible to record protons of kinetic energies up to 40–100 Mev. For an emulsion of a given type the limit is set by the lowest value of the ionization required to leave recognizable tracks. This value of ionization is, however, still much greater than the minimum ionization which can be produced by a charged particle. For a particle of given charge the minimum ionization occurs when the velocity of the particle approaches c or, in terms of energy, when the energy of a particle is approximately twice its rest energy. For an electron this means an energy of about 1·0 Mev., for a π-meson 300 Mev., and for a proton 2,000 Mev. The ionizations produced by protons of kinetic energies 40 Mev. and 100 Mev. are 7 and 3·5 times the minimum ionization.

It is clear that in emulsions similar to the Ilford C2 emulsion all the evaporation particles are recorded. Many of the knock-on particles are, however, lost. The recording of these particles, and indeed of all particles with velocities approaching c, has been made possible by the discovery of new emulsions by Messrs. Kodak and Ilford. These are called electron-sensitive emulsions, and the commonest types in use are designated NT4 (Kodak) and G5 (Ilford). It has thus now become possible to record all the particles emitted from stars. A typical set of results on the new Kodak plates obtained by Brown and her collaborators (1949) at an altitude of 3,450 m. is shown in Figure 7. In this diagram, which refers to stars with seven or more heavily ionizing particles, the numbers of particles are plotted as a function of the number of grains per millimetre of track. The number of grains per millimetre corresponding to minimum ionization is between 300 and

420, depending on the particular batch of plates. It is seen that the curve has two peaks, one for ionizations of approximately 1 to 1·5 times minimum and another for a value of the ionization greater than 5 times minimum. Between these peaks there is a continuous band of particles. These empirical groups of tracks are called 'thin', 'black' and 'grey' respectively. Clearly the black tracks are just the evaporated particles considered already: the grey tracks are mainly knock-on protons and the thin tracks are created particles. Fowler (1950) has given convincing proof that this classification is correct. In a recent paper, which will be considered later, he has shown that the grey tracks are mainly protons and that the thin tracks are mainly π-mesons.

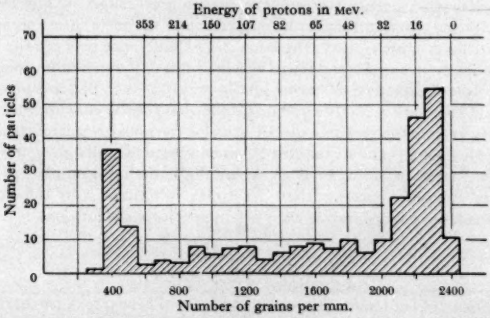


Figure 7. Histogram showing a typical distribution in the values of the grain-density in the tracks associated with stars. The measurements are confined to tracks of length greater than 300 μ and refer to stars of seven or more prongs. (*After* Brown *et al.* 1949)

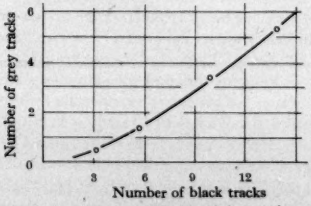


Figure 8. Number of grey tracks as a function of the number of black tracks for stars of seven or more prongs. (*After* Brown *et al.* 1949)

Brown *et al.* (1949) find that the black tracks are distributed isotropically; the grey tracks, as might be expected, show a tendency for collimation downwards.

We consider now the relation of the grey tracks to the black tracks.

The mean excitation energy due to a fast nucleon traversing a heavy nucleus is expected theoretically to be about 100 MeV. Thus if the mean observed excitation energy of stars is greater than 100 MeV. it is probable that the additional energy comes from the fast secondary nucleons knocked-on by the primary. Now the average excitation energy is about 200 MeV. and, therefore, the contribution of the knock-on nucleons must be appreciable. The average number of grey tracks as a

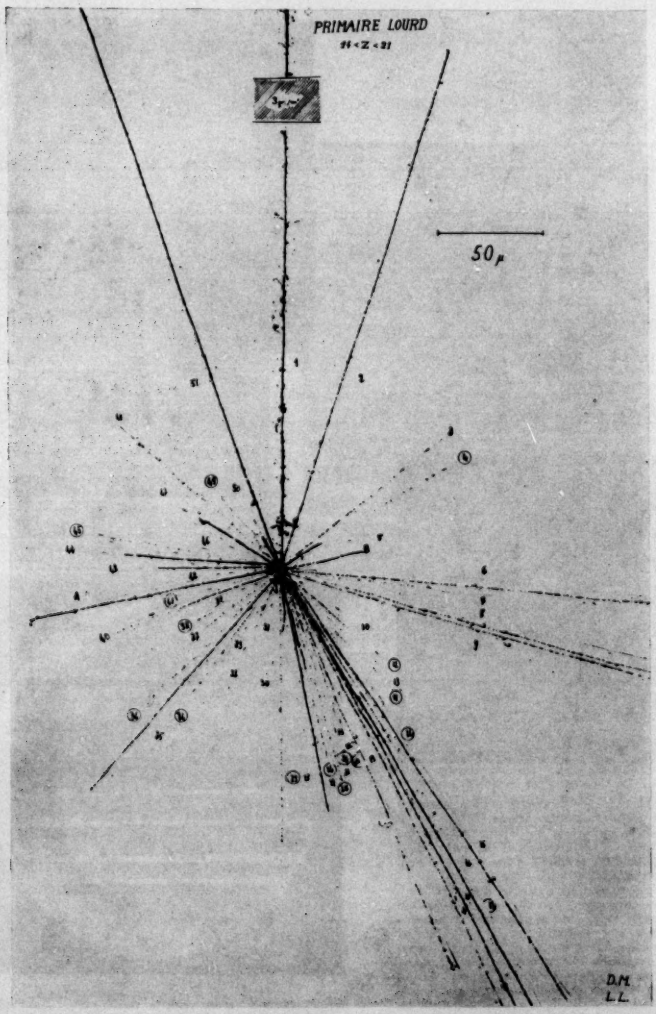
function of the number of black tracks for stars of gradually increasing size is shown in Figure 8, which has been drawn from the data of Brown *et al.* It is seen that the ratio of grey to black tracks is fairly constant for all star sizes and is about 1 : 3. If it is assumed that neutrons and protons are knocked out with equal probability in the 'grey' region, and if we neglect the small neutron excess, the average number of evaporation prongs per knock-on nucleon is about 1·5. This figure is confirmed by comparing the average size of stars produced by deuterons with energies of about 100 Mev. from the Berkeley cyclotron. In this case both a proton and a neutron in the grey region hit the nucleus. As already indicated in § 1.3, the average number of evaporation prongs is about three, which again means about 1·5 evaporation prongs per nucleon in the grey range. Thus for both cosmic-ray induced stars and deuteron induced stars the ratio of evaporation to knock-on particles is approximately the same. The difference between the energy loss of a fast nucleon in crossing a nucleus and the value of the excitation energy can therefore be fully understood in terms of the grey tracks. Bernardini and his collaborators (1950) have recently arrived at somewhat similar conclusions.

The total energy in the grey and black tracks for stars of size greater than seven prongs (grey plus black) can be obtained by an empirical formula given by Brown *et al.*, namely $E\,(\text{Mev.}) = 37\,N_H + 4\,N_H^2$, where N_H is the total number of tracks observed with ionization greater than 1·5 times minimum. In this formula allowance is made for energy carried away by α-particles and neutrons.

1.5. *Heavy Fragments*

Careful examination of the centre of a star shows that in many cases there exist short thick tracks of length between 3 and 6 μ. These tracks are ascribed to the nucleus recoiling during and after the impact of the fast particle producing the star. An extensive study of these nuclear recoils has been made by Harding (1949a). He finds that 34% of stars of size 2–5 prongs and 58% of stars of size 6–19 prongs have visible recoil particles. From the mean range he has deduced that the mean momentum transfer in heavy nuclei is about 175 Mev/c. He estimates the mean excitation energy as 190 Mev. Thus the energy transfer is about $\boldsymbol{c}$ times the momentum transfer. A neutron which excites a nucleus without being captured loses momentum in excess of $1/\boldsymbol{c}$ times the energy transfer except in the limit when the velocity of the neutron approaches $\boldsymbol{c}$. Harding's results are therefore consistent with the assumption that cosmic-rays stars are caused by fast neutrons which lose part of their energy while traversing the nucleus. Harding estimates that the mean velocity of recoil particles is approximately $0{\cdot}01\boldsymbol{c}$, whereas the mean velocity of the evaporated particles is about $0{\cdot}2\boldsymbol{c}$. Thus the recoiling nucleus cannot modify the angular distribution of evaporated particles in the laboratory frame of reference.

Occasionally fragments are emitted from stars with masses and charges much larger than α-particles. An example of the emission of a $^8_3$Li nucleus (marked Li) from a star S is shown in Plate II (Cosyns, Dilworth and Occhialini, private communication). The $^8_3$Li nucleus is unstable and decays into an excited $^8_4$Be nucleus with the emission of an electron. The excited $^8_4$Be nucleus then breaks up into two α-particles which fly off in opposite directions, producing the so-called 'Hammer Track.' The photographic emulsion used for this event was electron-sensitive and the electron can clearly be seen.

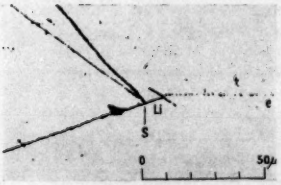


A large star, consisting of 34 heavily ionizing particles and 17 relativistic particles, produced by the collision of a heavy primary particle (charge 14–20) with a heavy nucleus in an electron-sensitive photographic emulsion. The star consists mainly of evaporation particles and has very few mesons. There are so many particles that the star must correspond almost to the complete disintegration of the incident and struck nuclei.

(*After* Leprince-Ringuet, Bousser, Hoang-Tchang-Fong, Jauneau and Morellet 1949)

PLATE I.

To face page 240

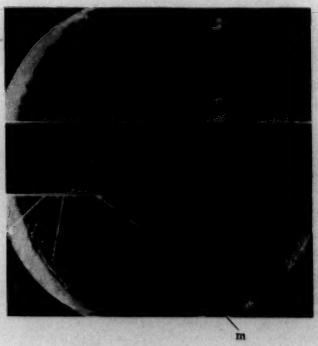


A hammer track.
(*After* Cosyns, Dilworth and Occhialini, private communication)

PLATE II.

A cloud-chamber photograph of a mixed penetrating shower which consists of a large electronic shower and four lightly ionizing penetrating particles. The penetrating particles are indicated by numbers. They are positively charged and have momenta ranging from 0·65 to $1{\cdot}4\times10^{8}$ ev/c. A positive particle coming down is curved in a clockwise direction. $H=7{,}200$ gauss.
(*After* Rochester and Butler 1948)

PLATE IV.



An example of meson production in lead obtained with a cloud chamber triggered by a set of counters arranged to select penetrating showers. The thickness of the lead plate is 3 cm. and the strength of the magnetic field 3,400 gauss.

The meson, indicated by the letter 'm', is a negative particle of momentum $2{\cdot}2\times10^{7}$ ev/c., and is produced in a high-energy star initiated by a fast charged particle entering the top of the plate. From the ionization and curvature the mass of the meson appears to be about 200 m_e. However, because of the difficulty of estimating ionization accurately, a π-meson cannot be excluded. A heavier particle is definitely excluded.
(*After* Rochester, Butler and Runcorn 1947)

PLATE III.

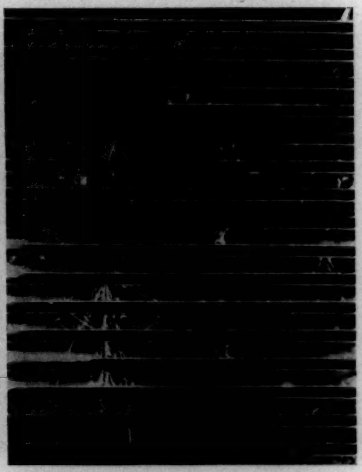
A photograph taken with a multi-plate cloud chamber showing the production of a large secondary nuclear interaction in a penetrating shower.
(*After* Lovati, Mura, Salvini and Tagliaferri 1950)

PLATE VIII.

A cloud-chamber photograph of a mixed shower which is produced in the glass plate above the chamber. The chamber contains lead plates (thickness 0·25 in.) and aluminium plates (thickness 0·313 in.) arranged alternatively. The top and bottom plates in the chamber are lead. The shower contains 7 or 8 penetrating particles, a high-energy electron shower which appears below the third lead plate, and several low-energy electron showers. A low-energy star occurs in the top plate.

(*After* Gregory and Tinlot)

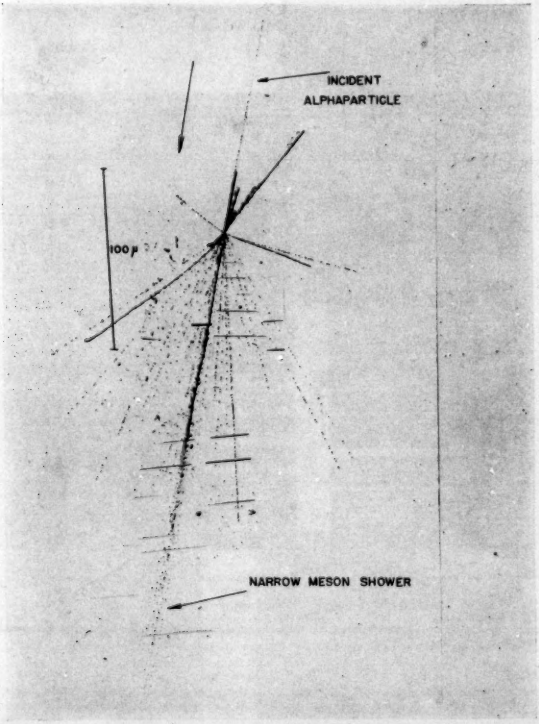
PLATE V.



An example of the successive production of penetrating particles in a penetrating shower. The cloud chamber contains 16 lead plates each of thickness 0·5 inch, giving a total thickness of lead in the chamber of 8 inches. The particle producing the shower penetrates the first three plates without interaction. Its ionization becomes noticeably heavier under the fourth plate, probably due to the creation of another particle. At the fifth plate a small star is produced which contains at least two heavily ionizing particles. At the tenth plate a larger penetrating shower is produced which contains at least five penetrating particles emitted in a wide core. Further down, a small electron shower occurs, and one of the penetrating particles, which is certainly a meson, slows down and stops.

(*After* Fretter 1949)

PLATE VII.



A meson shower produced by the collision of an extremely energetic primary cosmic-ray α-particle with a heavy nucleus, Ag or Br, in an electron-sensitive Kodak NTB3 emulsion exposed at an altitude of 100,000 ft. The collision gives rise to a very narrow core of some 23 relativistic, singly charged particles and a wider core of 33 relativistic particles. There are also 18 heavily ionizing particles carrying at least 23 units of charge. The energy of the primary α-particle is estimated at 10^{13} ev.

The narrow core of mesons has been followed into another stack of plates, and it is found that there is an appreciable increase in the number of relativistic particles due to the production of pairs of electrons. It is concluded that these come from photons produced by the decay in flight of π^0-mesons which are created in the original collision. The mass of the meson and its estimated lifetime agree with the results of the Berkeley experiments on high-energy γ-rays produced in nucleon–nucleon collisions.

(*After* Kaplon, Peters and Bradt 1949)

PLATE VI.

Investigations of heavy fragments have been made by Hodgson and Perkins (1949), Bonetti and Dilworth (1949) and Sorensen (1949). Perkins has found that particles with charges up to $Z = 10$ and with energies of 300 Mev. are emitted from stars. Approximately 1% of all stars seem to have long-range nuclear fragments, and the frequency of occurrence increases rapidly with star size. The mode of emission of these particles is not known but it is possible that some of them are fission products.

§ 2. PENETRATING SHOWERS

2.1. *The Investigation of Penetrating Showers by Counter Arrangements and the Cloud Chamber*

A penetrating shower may be defined as a high-energy interaction in which mesons are produced or, in terms of the classification introduced in § 1.4, a star in which thin tracks arise. In this section we shall survey the main results which have come from counter and cloud chamber investigations of such showers, choosing in each case only typical experiments to illustrate the main advances of the subject. Fuller accounts of the early work have been given by Jánossy (1948), Montgomery (1949) and Rossi (1948, 1949).

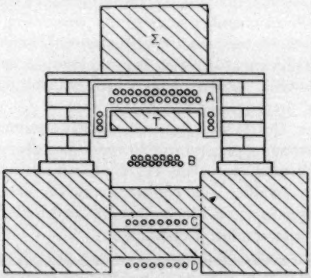


Figure 9. The arrangement used by Jánossy and Rochester for observing penetrating showers.

Probably the first evidence for the existence of a penetrating shower was a cloud-chamber photograph by Fussel in 1937 which showed a star from which came several contemporary particles capable of penetrating several lead plates without multiplication or appreciable deviation. The next main advance was made independently by Wataghin, de Santos and Pompeia (1940) and by Jánossy and Ingleby (1940) using various arrangements of counters embedded in large thicknesses of lead absorber. Jánossy and Ingleby, in particular, made a very careful analysis of the different events which would simulate penetrating showers, and they devised a very rigid selector for such showers (see also Jánossy 1942). A modified form of this shower selector is shown in Figure 9 (Jánossy and Rochester 1943). The penetrating shower is recorded by sets of counters B, C, and D, arranged in sevenfold coincidence (threefold at B and twofold at each of C and D). The showers recorded by this set are produced in the absorber T or come from the air. Completely surrounding the top and sides of T is a set of anticoincidence counters A the purpose of which is to find out which of the showers produced in T are due to

neutral and which to ionizing particles. Above A is placed another absorber Σ which is used to investigate further the nature of the shower-producing radiation and to find out if this radiation is coming into the apparatus from above. The main results of this investigation, and other similar investigations, are as follows. Penetrating showers are produced by both neutral and ionizing radiations of approximately equal intensities (see also Walker 1950). The rate of the showers produced in T increases steadily with the thickness up to a flat maximum at about 10 cm. lead. The producing radiation is coming downwards, and at sea level is almost completely absorbed in 40 cm. lead. The neutral part of the radiation does not consist of photons. The rate of such showers at sea level is about 1 per 10,000 μ-mesons or, with the arrangement shown in Figure 9, about one shower in two hours. Approximately one-third of the showers are connected with extensive showers.

The fact that the showers are produced by both neutral and ionizing radiation and that the neutral radiation is not photonic suggests that it consists of fast neutrons and protons (Jánossy 1943). This conclusion is in line with later cloud-chamber work, for example by Hazen and W. M. Powell, and experiments on the variation of penetrating showers with altitude, for example by George and Jason (1948).

The first detailed cloud-chamber study of the penetrating showers selected by a rigid counter arrangement was made by Rochester (1946). In this work clear proof was given that penetrating showers do contain ionizing penetrating particles. Photographs were obtained showing that some penetrating showers had a large electronic component, the origin of which could not be accounted for at the time. A year later Rochester, Butler and Runcorn (1947) obtained direct evidence for the production of mesons in penetrating showers (see also W. M. Powell 1946). This photograph is reproduced in Plate III, the meson being marked 'm'.

Since 1948 cloud chamber and counter work on penetrating showers has continued with investigations of the types and energies of the particles in the showers and of their interactions with matter.

It has already been stated that much evidence was accumulated in the early work showing the presence of mesons in penetrating showers. It was, however, not certain whether the mesons were π- or μ-mesons. This important point was cleared up by an ingenious counter experiment by Piccioni (1950) who showed conclusively that most of the slow mesons are π-mesons. He also deduced that a large fraction of the fast mesons must be π-mesons (see Powell 1950). Taken together with the experiments at Berkeley, which have shown the direct production of π-mesons by the collision of fast protons and α-particles with nuclei, Piccioni's experiment confirms the assumption that cosmic-ray mesons are produced by the collisions of fast nucleons with matter.

2.2. *Penetrating Showers in Electron-Sensitive Emulsions*

With the discovery of the electron-sensitive emulsion it has become possible to make an exact analysis of star types at any altitude and, by a study of the grain density and scattering along the tracks, to measure the energies and identify many of the fast particles. The results now to be described, which are largely the work of the Bristol School, represent one of the most significant contributions in the history of the subject.

Star types will be described by the method introduced by Brown *et al.* (1949). In their notation a star is designated by the number of heavy tracks N_H (black + grey), and the number of thin tracks n_s (called shower particles) emerging from it. Production of the star by charged or neutral particles is indicated by adding either the letter p or the letter n to n_s. Examples of different types are shown in Figure 10. The first star in this Figure, which has four heavy tracks

Figure 10. Schematic representation of stars of different types. (*After* Brown *et. al* 1949)

and four thin tracks, of which three thin tracks emerge in a narrow core, is designated 4 + 3p. Similarly, the second star, which has the same number of heavy tracks and five thin tracks, is designated 4 + 5n.

In Table 1 are given the results of Brown and her collaborators on the classification of the tracks from 4,923 stars formed in Kodak NT4 plates exposed at

Table 1. The Relative Frequency of Production of Stars at an Altitude of 3,450 m., in which a given number of shower particles n_s is associated with the emission of a number N_H of more heavily ionizing particles (grey and black tracks). (*After* Brown *et al.* 1949)

Number of heavily ionizing particles, N_H.

Number of emitted 'shower' particles, n_s.	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21–26	>27	Total
0^n	1030	1063	741	357	186	103	65	50	37	24	23	17	12	4	4	3	2	4	4	..	3729
0^p	41	61	92	91	71	54	46	18	9	13	6	5	6	5	4	1	1	0	2	..	525
1^n	24	33	37	33	19	17	18	14	8	8	4	11	2	3	1	3		2	1	..	238
1^p	32	24	34	26	19	10	11	13	4	8	5	6	5	1	1	2	1	..	4	2	208
2^n	5	8	8	6	5	4	5	5	3	2	..	2	1	..	1	1	..	..	..	..	54
2^p	2	7	6	6	7	6	5	3	4	5	5	2	..	1	1	3	1	..	..	..	69
3^n	1	2	2	1	1	.	3	2	..	..	..	..	..	1	.	2	1	..	..	..	16
3^p	1	2	4	..	4	..	..	..	..	4	1	1	1	1	1	1	..	..	..	..	21
4^n	..	..	..	2	1	..	..	1	.	2	..	..	..	1	1	1	1	..	..	1	11
4^p	1	..	1	1	1	1	..	1	..	1	1	2	..	..	1	2	..	..	..	..	13
5^n	..	..	..	..	1	..	..	1	1	..	..	..	..	..	..	..	..	..	..	..	3
5^p	..	..	..	1	2	..	1	..	..	2	1	..	..	..	..	..	..	..	..	..	7
6^n	..	1	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	2
6^p	1	..	1	..	..	1	..	..	..	..	..	..	1	..	1	..	..	1	..	..	6
7^n	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	0
7^p	..	1	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	1
8^n	..	1	..	..	..	..	..	..	..	..	..	..	..	1	..	..	..	..	..	..	2
8^p	..	..	..	1	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	1
9^n	..	..	..	..	..	..	..	1	..	..	..	..	..	..	..	..	..	..	1	..	2
9^p	..	..	..	..	..	..	1	1	..	..	..	..	..	..	..	..	..	..	..	..	2
10^n	..	..	..	..	..	..	1	..	..	..	..	..	..	..	..	..	..	..	..	..	1
10^p	..	..	..	1	..	..	2	..	..	..	1	..	..	..	..	..	..	..	..	..	4
11^n	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	0
11^p	..	..	..	..	..	..	..	1	..	..	..	..	..	..	1	..	..	..	..	..	2
12^n	..	..	..	..	..	..	2	..	..	..	..	..	..	..	..	..	..	..	1	..	3
12^p	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	0
13^n	..	..	..	..	..	..	..	..	..	..	..	..	1	..	..	..	..	..	..	..	1
13^p	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	↓ 17p	↓ 21n	..	2

To reduce the observed values to numbers per cm² per day, multiply by $\sim 2{\cdot}2 \times 10^{-3}$. The resulting numbers then give the estimated intensities of stars of different types in conditions in which there is no absorbing material above the plates.

an altitude of 3,450 m. Most of the results refer to plates kept under very thin absorbers, but a few of the plates were kept under 7·5 cm. lead. The rate of all stars with $N_H \geqslant 3$ is 10·8 stars/cm³/day. At the same altitude Cortini and Manfredini (1949) found $14{\cdot}2 \pm 0{\cdot}5$ stars/cm³/day on Ilford C2 plates, a value which has since been confirmed by several other workers. Thus the Bristol rate seems to be too low. The difference is unlikely to be due to fading, since little fading occurs in electron-sensitive plates; rather it is probably due, as the Italian workers have pointed out, to the severe loss of three- and four-prong stars by fast scanning. The correct rate is probably given by doubling the numbers of stars in Table 1 with $N_H = 3$ and 4. There is good agreement among all workers for the rate of the bigger stars. Approximately 25% of all the stars formed at 3,450 m. have thin tracks associated with them.

The thin tracks produced by charged particles are strongly collimated in the direction of the primary. Moreover, stars with $n_s \geqslant 2$ are produced in almost equal numbers by charged and uncharged particles. This result, which is in good agreement with the results of counter and cloud-chamber work on penetrating showers, strongly suggests that such stars are identical with penetrating showers. Confirmation comes from the agreement between the size–frequency distributions for shower particles from stars and penetrating showers observed in the cloud chamber at the same altitude (Fretter 1949 a).

The first extensive identification of fast star particles (grey and thin tracks) was made by Fowler (1950) by observations on the grain density and scattering along the tracks (see also Occhialini 1949, Camerini *et al.* 1949 and Goldschmidt-Clermont 1950). Such identification is possible because, for particles of given charge, a unique relation exists between grain density and scattering for each value of the mass. By plotting the logarithm of the number of grains per 50 μ against the logarithm of the mean deviation $\bar{\alpha}$ of the track per 100 μ in the manner shown in Figure 11, curves for the different star particles are obtained which closely resemble the familiar specific ionization–momentum curves already widely used for the identification of elementary particles in the cloud chamber. The experimental data shown in the Figure refer to 1,500 primary and secondary star particles of minimum length 3 mm. observed in G5 emulsions exposed at an altitude of 70,000 ft. by Camerini, Fowler, Lock and Muirhead (1950). The sensitivity of the method is very striking and the rapid methods introduced by Fowler (1950), particularly for the determination of $\bar{\alpha}$, have made possible the accumulation of a large mass of data. It is seen that the observed points separate into several well-marked mass and charge groups.

The distribution of particles at and above the mass 1,837 m_0 indicates the presence of many protons and smaller numbers of deuterons and tritons. The energies of all of these particles are so high that they must be either knock-on particles or primaries. There are also a few α-particles, all of which are primaries.

The particles with kinetic energies less than 150 MEV. are almost entirely π-mesons. The nature of the particles with kinetic energies in the range 150–1,400 MEV. which have an ionization less than 1·3 times minimum cannot be determined with certainty. Clearly these particles cannot be protons, but, from their apparent connection with the identifiable π-meson group, are almost certainly mesons.

Very few of the shower particles (i.e. thin tracks) can be electrons. This follows from the fact that although 20 of 300 shower particles have path lengths in the

emulsion greater than 2 cm., no significant decrease in energy has been observed for any particle. The radiation length in the emulsion is 2·3 cm. (Carlson *et al.* 1950).

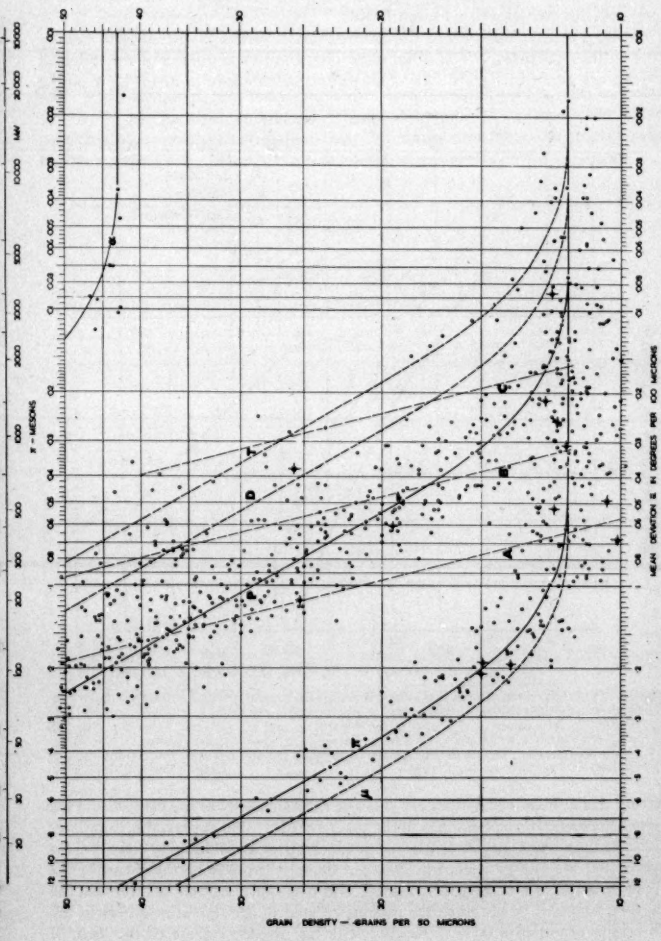


Figure 11. Relation between grain density and mean angle of scattering δ^0 per 100 μ cell length for the tracks of some 1,500 particles from stars obtained on plates exposed at altitudes greater than 70,000 ft. ● 'primary' particles, ○ 'secondary' particles, ✚ interconnecting particles. The lines marked μ, π, P, D, T, and α refer to μ-mesons, π-mesons, protons, deuterons, tritons and α-particles. A, B, and C are lines of constant momentum corresponding to 500, 1,000 and 2,000 Mev/c. respectively. The scales at the top of the figure give the kinetic energies of π-mesons and protons.
(*After* Camerini *et al.* 1950)

Again, the observations show that very few mesons intermediate mass (i.e. in the mass range 300–1,200 m_e) are produced in nuclear explosions of the types observed here. Fowler estimates that the number must be less than 2% of the π-mesons. This result does not conflict with recent results on the existence of heavy unstable particles (see Powell 1950).

Fowler's results also permit the determination of the distribution in energy of the emitted particles. The differential distributions in energy of the different particles are given in Figure 12. The proton distribution is given by a power

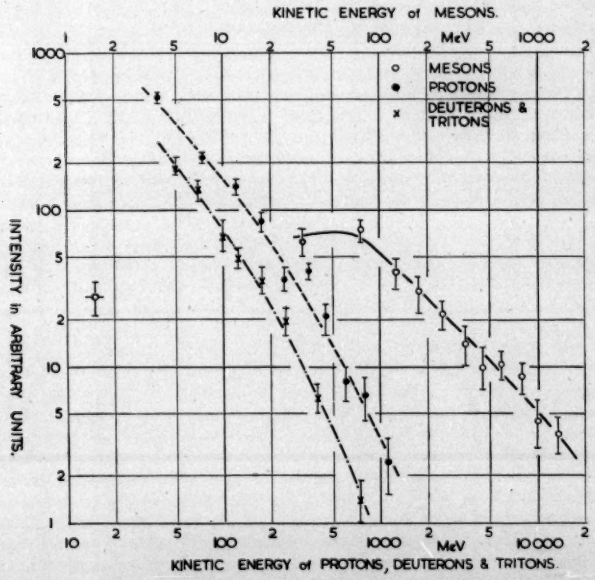


Figure 12. The differential distribution in kinetic energy for mesons, protons, deuterons and tritons from stars formed in the electron-sensitive photographic emulsion at altitudes greater than 70,000 ft. The point at 1·6 MeV. is for π-particles and is calculated from the observed number of mesons which are ejected from stars and produce nuclear disintegrations at the end of their range. To obtain the number of particles per star per MeV. interval multiply by $2{\cdot}5\times10^{-5}$. (*After* Camerini *et al.* 1950)

spectrum of the form $dE/E^{2\cdot5}$, where E is the kinetic energy of the proton, and the π-meson spectrum, by a power law of the form $dE/E^{1\cdot5}$. No appreciable difference is found in the energy spectra of mesons produced in stars of different sizes. This suggests that the energy spectrum does not depend strongly on the nature of the nucleus in which the mesons are created, and it is therefore fair to infer that the same spectrum will be obtained for mesons created in the free atmosphere at the main meson production layer. Sands (1950) has already calculated the form of the energy spectrum of μ-mesons at the production layer. Camerini *et al.* (1950)

have modified his spectrum by taking into account $\pi \rightarrow \mu$ decay and the difference in mass between the π-meson and the μ-meson, and they obtain a curve which closely follows the experimental points and gives a power law of the observed form.

At kinetic energies greater than $10\,mc^2$, where m is the rest-mass of the π-meson, Sands' results give a distribution of the form $dE/E^{2\cdot5}$. This formula can be used for the extrapolation of the meson spectrum beyond 1,400 MEV., and in conjunction with the observational data can be used to give the relative numbers of the various particles in different energy ranges. With this assumption it is found that of tracks with grain densities less than four times minimum $36 \pm 2{\cdot}5\%$ are π-mesons, $51 \pm 2{\cdot}9\%$ are protons, and $13 \pm 1{\cdot}5\%$ are deuterons or tritons. For shower particles with grain densities less than 1·3 times minimum the corresponding figures are $79 \pm 6\%$, $17 \pm 3\%$ and $4{\cdot}0 \pm 1{\cdot}3\%$ respectively.

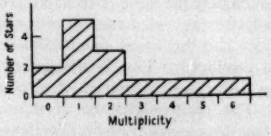


Figure 13. The frequency–shower-particle multiplicity distribution for primary particles (which may be protons or π-mesons) in the energy range 4,000–5,700 MEV.
(*After* Camerini *et al.* private communication)

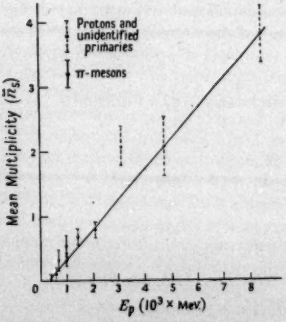


Figure 14. Mean shower-particle multiplicity ($\bar{n}_s$) as a function of the energy of the primary ($\bar{E}_p$) in 10^3 MEV. The broken lines refer to proton and unidentified primaries and the continuous lines to π-meson primaries.
(*After* Camerini *et al.* private communication)

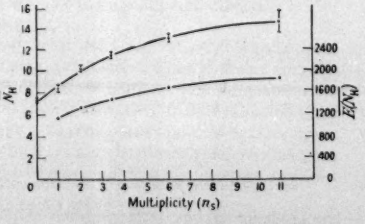


Figure 15. Shower-particle multiplicity as a function of the mean number of grey and black tracks $\bar{N}_H$ and the mean energy of grey and black tracks $\bar{E}(N_H)$.
(*After* Camerini *et al.* private communication)

The work of Camerini *et al.* (1950) has recently been extended by Camerini, Fowler, Lock, Muirhead and Yekutieli (private communication) to the determination of the energies of the primary particles which produce meson showers of different sizes. Protons can be distinguished from other particles of charge e when their kinetic energies are below 850 MEV., and π-mesons when their kinetic energies are below 1,100 MEV. Some typical results are shown in Figures 13–15. In

Figure 13 is shown the spread in multiplicity of shower particles for primaries (which may be π-mesons or protons) in the energy range 4,000–5,700 MeV. The wide spread of multiplicity is at once evident. In Figure 14 there is plotted the mean multiplicity of shower particles as a function of the energy of the primary particle. Identified π-mesons and protons are shown. Finally, in Figure 15 are plotted the mean number $\overline{N_H}$ and mean energies of black and grey tracks as a function of multiplicity. The slow rise in $\overline{N_H}$ and $\overline{E(N_H)}$ with production of shower particles is very striking. The same variation is found for grey tracks. Now it has already been noted in § 1.4, Part II, that the excitation energy, i.e. the energy of the black tracks, can be accounted for in terms of the energy lost by the primary particle and the knock-on nucleons in heating up the nucleus. This fact is shown by the constancy of the ratio of black to grey tracks whatever the size of the star. Approximately the same ratio is found in meson showers, which means that the evaporation energy of a meson star still comes almost entirely from the primary particles and the knock-on nucleons, and changes little with the number of mesons produced. This is a surprising result, for relativistic mesons can produce stars with a cross section equal to the geometrical cross section (see § 3.5, Part III). The relative absence of meson interaction may be a consequence of the fact that mesons are often produced in the form of a *very* narrow cone (see, for example, Plate VI).

Camerini *et al.* (private communication) have also obtained some interesting data on the energies of shower particles. For a given multiplicity (value of n_s) the mean energy of the shower particles tends to decrease slowly with increasing N_H, and for a given value of N_H the mean energy of the shower particles tends to increase slowly with n_s (Table 2).

Table 2. The Energies of Shower Particles (MeV.) as a Function of N_H and n_s

n_s \ N_H	3–8	9–14	15–47
1	725±80	500±80	580±90
2–4	725±80	770±120	800±120
5–8	1040±100	790±120	610±90

2.3. *The Determination of the Momentum Spectrum of Shower Particles by the Cloud Chamber*

Finally, it may be noted that the counter-controlled chamber has been applied at sea level to the determination of the momentum spectrum of some of the shower particles which can penetrate a few centimetres of lead. Most of the observed particles arise in showers formed in the lead outside the chamber, and it is difficult to estimate how much the spectrum will be biased by the complicated geometrical disposition of counters and absorbers around the chamber. Geometrical effects are likely to be less important at high momenta, and for this reason the results to be presented now, which are taken from a summary by Barker and Butler (1951), refer only to momenta above 1,000 MeV/c. Eighty-five particles are observed in this region of the spectrum, and they fit a differential spectrum of the form $dp/p^{(2\cdot5\pm0\cdot3)}$. The positive excess of particles in the momentum range 1,000–2,000 MeV/c. is $2\cdot9\pm0\cdot7$, a value in good agreement with the corresponding figure found by

Camerini *et al.*, namely $2{\cdot}4 \pm 0{\cdot}9$. Above 2,000 Mev/c. the positive excess is $2{\cdot}5 \pm 0{\cdot}8$, a figure which would seem to indicate that even at high momenta an appreciable fraction (perhaps 25%) of the shower particles are protons.

2.4. *Mixed Showers*

Mixed showers may be defined as showers containing both soft and hard components, that is, electrons (and photons) and penetrating particles. Evidence for the presence of soft component in a shower selected as a penetrating (or hard) shower has been given in § 2.1; the converse, namely the observation of hard component in showers selected as soft showers (for example, extensive showers), has also been clearly shown (Jánossy 1942, Daudin 1942, 1943, Cocconi *et al.* 1946, Broadbent and Jánossy 1947). Apart from the recent work with the photographic emulsion, the investigation of these mixed showers has proceeded along three main lines: (i) penetrating showers, (ii) bursts, and (iii) extensive showers. The results of this work, which will now be discussed, are to demonstrate the intimate connection which exists between the soft and hard components.

(i) The early work on penetrating showers was extended by Fretter (1948) at mountain altitudes using a counter-controlled cloud chamber containing several lead plates. Fretter obtained many photographs of nuclear interactions. Fretter's work was confirmed and extended by Chao (1949) who also used a cloud chamber containing lead plates. Chao noticed that the electron showers were often seen to be connected with the main interaction by non-ionizing links and that sometimes small showers diverged at large angles from the main shower direction. Direct evidence for the identification of these shower particles as electrons was obtained by Rochester and Butler (1948), an example of whose work is shown in Plate IV, and Butler, Rosser and Barker (1950). These authors, and also Lovati, Mura, Salvini and Tagliaferri (1950), and Brown and McKay (1950) obtained many examples of mixed showers.

A recent example of a mixed shower, taken by Gregory and Tinlot (private communication) is shown in Plate V. These authors use a cloud chamber containing seven 0·25 in. lead plates placed alternately with aluminium plates each 0·313 in. in thickness. Each aluminium plate is equivalent to 0·08 radiation units and each lead plate to 1·2 radiation units. Examples are found of nuclear interactions in the aluminium plates, with which are associated small electron cascades in the lead plates. Non-ionizing links frequently connect the electron cascades with the main nuclear interaction.

(ii) The work on bursts was carried out mainly by Rossi and his collaborators (Bridge, Rossi and Williams 1947, Bridge and Hazen 1948, Bridge and Rossi 1949). A typical arrangement, shown in Figure 13, consists of an ionization chamber triggered by counters. Between the tray and the chamber is a large thickness of dense absorber. Bursts recorded by the ionization chamber may be due to the following events: (*a*) electron showers from the air, (*b*) collision or radiation showers from μ-mesons, (*c*) nuclear interactions, (*d*) heavy primary cosmic-ray particles. The relative number of each type of event which is observed depends on the altitude, the nature of the absorber surrounding the chamber, the dimensions of the chamber and the nature and pressure of the gas inside the chamber. With a shielded chamber at mountain altitudes events (*a*) and (*d*) are almost completely negligible. Moreover, as will be shown in a later section, the contribution of event (*b*) is unimportant except at sea level. We are left with events of type (*c*).

Ionization bursts of this type can be produced in three ways : (i) by the passage through the chamber of a number of lightly ionizing particles, (ii) by the passage of one or more heavily ionizing particles and (iii) by associated events of types (i) and (ii). A very careful study of the shapes of the pulses recorded by the chamber has permitted the separation of events (i) and (ii), and Rossi and his collaborators have shown that the shielded ionization chamber selects *electron* showers produced in the lead absorber. It will be seen later (§ 5.3, Part III) that a completely different selection is made by an unshielded ionization chamber.

These authors have also demonstrated that this electronic component shows the same type of transition effect in the atmosphere as do penetrating showers. It is thus clearly connected with penetrating showers.

In some of the experiments pulses of different sizes were recorded separately. It was found that at large lead thicknesses the number $N(P)$ of pulses larger than a certain P decreases with increasing P according to the law $N(P) = \text{const.} \times P^{-1.5}$.

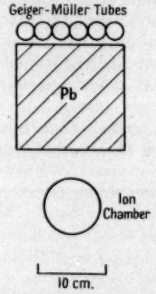


Figure 16. Arrangement of ionization chamber and G–M counters for the investigation of mixed showers. (*After* Bridge, Rossi and Williams 1947)

This result may be interpreted as meaning that the integral energy spectrum of the electrons or photons produced in nuclear interactions obeys a power law with the same exponent.

(iii) Finally, it may be noted that the large extensive showers in the atmosphere contain a hard component. It is beyond the scope of this review to give all the steps of the detailed analysis which led to this result. Suffice it to say that the evidence for the presence of an ionizing penetrating component of extensive showers is now quite convincing. The cloud-chamber work of Mitra and Rosser (1949), Ise and Fretter (1949) and Brown and McKay (1949) shows that the penetrating particles are mesons, and probably μ-mesons. Counter work by Chowdhuri (1950), Sitte (1950b) and McCusker (1950) shows that below 20 cm. of lead the penetrating component of extensive showers consists of some 80% mesons and 20% of nuclear-interacting particles.

It is thus abundantly clear that there exists a very close connection between the soft and hard components of showers. Only recently, through the fundamental experiments of Bjorkland, Crandall, Moyer, and York (1950), has the reason for this connection become plain. These workers found that pairs of electrons are

produced when protons of energy greater than 180 MEV., accelerated in the Berkeley cyclotron, fall on various absorbers. The pairs clearly arise from photons, and Bjorkland *et al.* made a careful study of the ways in which photons can be produced when matter is bombarded by high-energy particles. One suggestion, due originally to the Japanese physicist Hayakawa (1949), is that the photons are due to the deceleration of the protons. The cross section for this process is, however, not sufficient to account for the Berkeley results. Bjorkland and his collaborators came to the conclusion that the photons arise from the decay in flight of neutral mesons of mass about $300\,m_e$ which have a lifetime of less than 10^{-11} second. These neutral mesons are assumed to be among the primary products of the nuclear interaction.

Further experimental evidence supporting this conclusion has come from the star shown in Plate VI which was obtained by Kaplon, Peters and Bradt (1949). This star was found in a Kodak NTB3 electron-sensitive emulsion exposed at an altitude of 100,000 ft. The star has a narrow core of 23 shower particles which is surrounded by a wide core of 33 shower particles together with 18 evaporation particles. It is produced by an α-particle. The core of the shower can be followed into the next plate after passing through a thickness of glass equivalent to 0·26 radiation units. It is observed that the number of shower particles increases, the difference being due to the appearance of 8 electron pairs. The number of photons corresponding to this frequency of electron-pair production in so small a thickness of glass is much too large to be accounted for by bremsstrahlung, and is, in fact, comparable with the number of charged particles in the core. The photograph can be completely interpreted if the photons arise from the decay of neutral mesons which are produced with the charged mesons, and are equal in number to about half of the charged mesons. An upper limit for the lifetime of the neutral meson can be estimated roughly and is given as 3×10^{-13} second.

Recent experiments on the capture of negative π-mesons by liquid hydrogen by Panofsky, Aamodt and York (1950) seem to have established the existence of the neutral π-meson beyond doubt. The π^--meson is absorbed by the following process: $^1\mathrm{H} + \pi^- \rightarrow \mathrm{n}^0 + \pi^0$; $\pi^0 \rightarrow 2h\nu$, where π^0 represents the neutral meson. The π^0-mesons are emitted with small kinetic energies, and from the known masses of the other particles in the reaction lead to a mass of $\sim 280 m_e$.

Carlson, Hooper and King (1950) have observed the electron pairs produced by the γ-rays in Ilford G5 emulsions exposed at an altitude of 70,000 ft. The detailed form of the spectrum of the γ-rays is consistent with the assumption that the rays originate in neutral mesons presumably produced in large stars (see Powell 1950). The lifetime is shown to be less than 5×10^{-14} second and the fraction of neutral to charged mesons is found to be $0{\cdot}45 \pm 0{\cdot}10$ (cf. Kaplon, Peters and Bradt 1949).

Further evidence for the existence of π^0-mesons comes from a consideration of the balance of energy in stars by Camerini *et al.* (private communication). These authors have examined the balance for a number of stars with $n_s \leqslant 4$, including $n_s = 0$, which are formed by primary particles of mean energy $\bar{E}_p = 2,660 \pm 200$ MEV. Assuming two-thirds of the fast shower particles, i.e. those with a mean angle of scattering $\bar{\alpha} < 0{\cdot}0226°/100\,\mu$, to be $\pi^\pm$-mesons and the rest protons, and allowing a reasonable figure for the number of neutrons emitted, it is found that the total energy of the emitted particles is 2,195 MEV. The evaporation particles, charged and neutral, contribute 1,130 MEV. to this total, identified $\pi^\pm$-mesons 605 MEV.,

proton and neutron shower-particles 460 Mev. Thus the missing energy is 465 ± 200 Mev. If this energy has gone into neutral π-mesons with an energy spectrum equal to that of the $\pi^{\pm}$-mesons, the ratio $\pi^0/\pi^{\pm}$ is $0{\cdot}7 \pm 0{\cdot}3$, a value in reasonable agreement with other estimates.

2.5. *Heavy Unstable Particles in Penetrating Showers*

In 1947 Rochester and Butler observed two unusual events in cloud-chamber photographs of penetrating showers which they ascribed to the spontaneous decay in flight of an unstable neutral particle of minimum mass $(770 \pm 200)m_e$ and an unstable charged particle of minimum mass $(980 \pm 150)\,m_e$. These observations were confirmed by Seriff, Leighton, Hsaio, Cowan and Anderson (1950), who found 24 examples of the neutral particle and four of the charged particle at an altitude of 3,200m., and six examples of the neutral particle at sea level. Recently the cloud chamber and magnet used by Rochester and Butler in their original work have been set up on the Pic du Midi at an altitude of 9,500 ft., and Armenteros, Barker, Butler, Cachon and Chapman (1951) have found a further 36 examples of the neutral particle decay and five of the charged particle decay. These particles are now called V-particles, from the forked appearance of their tracks in the cloud chamber. Estimates of the ionization show that all the charged particles observed in the decay events have charges equal to the electronic charge.

Since the data on the charged V-particles are still very meagre only the neutral V-particles will be discussed. We will consider now the evidence for the nature of the two charged decay particles of the neutral V-particles. In no case can both of these particles be identified, but Seriff *et al.* find one case in which one of the secondary particles is of mesonic mass. Armenteros *et al.* have made measurements on 54 secondary particles of which 25 are positive and 29 are negative. Of the positive particles four are identified as protons, three are at minimum ionization and cannot be protons, and 18 are at minimum ionization and can be protons or mesons. Of the negative particles three are identified as mesons, 18 are at minimum ionization and must have masses less than the proton, and eight are at minimum ionization and can be protons or mesons. Seriff *et al.* and Armenteros *et al.* find no examples of secondary particles producing electron cascade showers in a lead plate in the chamber. This shows that the particles are not as light as electrons. It is concluded that the particles consist of protons and mesons, presumably π-mesons. Seriff *et al.* find 12 cases in which the main nuclear interaction lies in the plane of the V-track; it is therefore deduced that the decay scheme involves only two particles. If this is correct a possible decay scheme for cases in which a proton and a meson are produced is $V^0 \rightarrow {}^+p + \pi^-$. With this scheme the mass of the V^0-particle is about $2250m_e$.

Some of the positive secondary particles cannot be protons. If in these cases the negative particle is π^--meson the decay scheme may be $V^0 \rightarrow \pi^+ + \pi^-$. This scheme leads to a mass value of about $1{,}000m_e$. Thus the photographs provide strong evidence for two different sorts of neutral V-particle. This conclusion can only be avoided by assuming the existence of the negative proton.

Seriff *et al.* estimate the lifetime of the V^0-meson to be about 3×10^{-10} second and its frequency of occurrence in penetrating showers to be of the order of a few per cent of that of the penetrating particles in the shower.

It is still too soon to say what all these strange results mean. Much more experimental evidence will be needed to establish the masses of the various types of particle, to show the relation between them, and to account for their properties.

§ 3. THE NUCLEAR INTERACTIONS OF HEAVY NUCLEI

One of the most remarkable developments of the post-war era has been the discovery by groups of workers at Rochester and Minnesota Universities of the presence of heavy nuclei in the primary cosmic radiation. The methods of identification of these particles, the proof that they are primaries, and the main results are summarized in two important papers by Bradt and Peters (1949, 1950). By sending photographic plates up to altitudes of 90,000 ft. it is found that the primary cosmic-ray beam consists of protons, helium nuclei and heavier nuclei. Of the heavy primaries with $Z \geqslant 6$, 75% belong to the carbon, nitrogen and oxygen group, and of the nuclei with $Z \geqslant 10$, 60% belong to the magnesium–silicon group ($11 \leqslant Z \leqslant 15$) and 10% to the iron group. No nuclei heavier than iron have been reported. At a latitude of 30° N. the relative intensities of hydrogen, helium, carbon–nitrogen–oxygen and nuclei with $Z \geqslant 10$ is approximately 4,000 : 1,000 : 35 : 10. The total intensity of helium nuclei at this latitude is $I_0(\text{He } 30°) = (9 \pm 3)10^{-3}$ nuclei cm^{-2} sec^{-1} sterad^{-1} and at a latitude of 51° is $I_0(\text{He } 51°) = (38 \pm 13)\ 10^{-3}$ nuclei cm^{-2} sec^{-1} sterad^{-1}. The spectrum is similar at both latitudes. The remarkable fact emerges from these data that some 20% of the primary particles at latitudes 30° and 51° are helium nuclei. This means that more than half of the nucleons entering the earth's atmosphere do so in the form of nuclei heavier than protons. The average energy per nucleon seems to be about 4,000 Mev. at $\lambda = 50°$, a value just above the cut-off energy imposed by the earth's magnetic field. The differential energy spectrum of the heavy nuclei ($Z \geqslant 2$) has been shown by Vallarta (1950) to be approximately of the form $N(>E) \propto E^{-1\cdot6}$.

The collisions of these heavy nuclei with the nitrogen and oxygen nuclei in the air must lead to the formation of both charged and neutral π-mesons in just the same manner as in the Berkeley experiments with fast, artificially accelerated, α-particles. An example of the collision of a primary α-particle with a silver or bromine nucleus has already been discussed (Plate VI). In Plate I is shown the production of a star of 51 particles by a nucleus heavier than an α-particle. This particle, which is identified as having a value of Z between 14 and 20, has an energy of more than 12,000 Mev. (Leprince-Ringuet *et al.* 1949).

Nuclei heavier than α-particles seem to undergo three main types of collision, examples of which are shown in the papers of Bradt and Peters. In the first type the incoming nucleus may knock out a few nucleons from the target nucleus and proceed undeflected and unchanged. In the second type part of the bombarding nucleus may be broken off and a star result which contains a core consisting of the remaining nucleus together with fast α-particles and protons; in the third type of collision the incoming nucleus may be completely broken up into the constituent nucleons. The second two types of collision may also involve meson production. The division into these three groups is not well defined. The various types probably depend on the energy of the incoming nuclei and the areas of overlap with the target nucleus. The question of the collision cross sections of these heavy nuclei will be taken up in § 2, Part III.

§4. THE NUCLEAR INTERACTIONS OF THE NON-NUCLEONIC COMPONENTS OF COSMIC RAYS

4.1. *Electrons and Photons*

Since the soft component is, in general, such a large fraction of cosmic radiation, it is pertinent to enquire whether electrons and photons can give rise to evaporation stars and penetrating showers. This was once thought to be the case (see Bagge 1943) because of the close parallelism between the rate of increase of stars and the soft component with altitude. Moreover, stars and heavily ionizing particles are often seen associated with large cascade showers in the cloud chamber. Apart from these observations, which seem to suggest the production of stars by the soft component, there is much evidence to the contrary. Some of this evidence has been given already (§1.1). Other evidence comes from the study of star production by the γ-rays produced in betatrons and synchrotrons. Baldwin and Klaiber (1946) investigated star production in air with a cloud chamber irradiated with the γ-rays from a 100 MEV. betatron and they found seven flag tracks and three 4-prong stars in 1,500 exposures. A value for the cross section for the process was not given but since these stars are almost certainly produced by photo-disintegration the cross section must be between 10^{-27} and 10^{-29} cm^2 per atom. As the energy of the γ-ray is increased above 100 MEV. a value is reached at which meson production can take place. The cross section for this process is found to be 5×10^{-28} cm^2 per carbon nucleus at 335 MEV. (Peterson, McMillan and White 1950). Apart from photo-disintegration and interaction through the meson field, there are very few processes by which photons can get energy into the nucleus. The Compton effect for nucleons would be one possible method, but the cross section for this process must be very small. Thus the maximum cross section for star production by γ-rays would seem to be about 10^{-27} cm^2 per light nucleus. The cross sections for star production by electrons are likely to be 137 times smaller than for γ-rays. Now the value of 10^{-27} cm^2 per light nucleus is 1,000 times smaller than the cross section for pair production by γ-rays. Thus star production will not be able to compete with pair production. Confirmation that the soft component does not play an important part in star production comes from an examination of the variation in the frequency of stars with altitude in the atmosphere. It is found that star production does not follow the soft component near the top of the atmosphere, and in fact the maximum production of stars occurs above the maximum of the soft component.

Since evaporation stars are produced by a neutral component, it is concluded that this consists of neutrons. George (1949) estimates from star production that the neutron flux in the energy range 100–1,000 MEV. is 6% of the flux of charged particles at sea level, and 20% at an altitude of 3,450 m. These values are consistent with direct determinations of the neutron flux by Tongiorgi (1949). All observations on evaporation stars are in agreement with their origin in the neutron component. The observation of stars in cascade showers is not in disagreement with this hypothesis since these showers are probably examples of mixed showers (§2.4).

4.2. *μ-Mesons*

Experiments on the absorption of μ-mesons have shown that these particles have a very small interaction with matter, the cross section for nuclear interaction being probably as low as 10^{-30} cm^2 per nucleon. Thus very few stars should be

induced by μ-mesons. Recently, George and Evans (1951) have made the remarkable discovery that photographic plates exposed at a depth of 60 m. water-equivalent underground contain a small number of stars. The plates used for this purpose were Ilford C2 and G5 and, to avoid contamination, were coated underground. By scanning a very large volume of emulsion George and Evans were able to establish a star rate of 0·005 per cm^3 per day. This rate may be compared with the rates at sea level ($\sim$1 per cm^3 per day) and at 3,450 m. ($\sim$ 15 per cm^3 per day). The star types observed below ground seem somewhat similar to those observed at sea level. A careful analysis of the phenomenon, together with evidence from other underground experiments, excludes all sources other than μ-mesons. The stars found in the electron-sensitive emulsion are of two types, namely, those with fast charged primaries and those with neutral primaries. The former group have the appearance of penetrating showers: the latter are typical evaporation stars such as are produced by neutrons. Following a suggestion by Heisenberg, it is assumed that the former group arise by the interaction of the Coulomb field of the μ-meson with the meson field of the nucleus. By analogy with the Williams–Weizsäcker approach to the brems-strahlung of electrons the μ-meson is imagined as being accompanied by virtual photons. These virtual photons are assumed to interact with the nucleus through the meson field in the same way that real photons produced in the Berkeley synchrotron produce mesons. Using the experimental value of the cross section for the real photons, George and Evans find that the cross section for the production of stars by μ-mesons at 10^{10} ev. is about 10^{-29} cm^2 per nucleon. This value is in agreement with the absolute value of the number of stars observed and their variation with depth. George and Evans also calculate the probable contribution of this type of star to the observed star population at different altitudes. At an altitude of 3,450 m., for example, 0·05 stars per cm^3 per day would arise from this cause.

George and Evans also show that the stars with neutral primaries are secondary to the meson-induced stars via knock-on neutrons produced in these stars.

It is important to enquire what contribution to star rates can be made by slow negative μ-mesons. Slow positive μ-mesons always decay into positive electrons (see Powell 1950). It is unlikely that many stars will be produced directly by slow μ-mesons because of the small interaction of these particles with nuclei. In light materials, like carbon, nitrogen and oxygen, the negative μ-meson decays into a negative electron and probably two neutrinos. In heavy materials, like silver and bromine, capture is more probable than decay, and the rest energy of the μ-meson ($\sim$ 100 Mev.) must be liberated. From considerations of spin, since the spin of the μ-meson is believed to be one-half and the spin of the proton is one-half when the μ-meson is captured by a proton, a neutron and also another light neutral particle of spin one-half must be emitted. If this neutral particle is a neutrino, it will take a large fraction of the energy; moreover, since the neutrino has little interaction with matter, little energy will go into the excitation of the nucleus. Rosenbluth (1949) has shown that the average excitation energy to be expected when a μ-meson is captured is only 10–15 Mev., the maximum being 22 Mev. This energy is sufficient to produce a star consisting of one prong or at the most of two prongs. George and Evans (1951) have recently found such stars underground. They estimate that

$8{\cdot}7 \pm 1{\cdot}7\%$ of all the negative μ-mesons captured in the emulsion give visible stars. Of these stars 90% have one prong and 10% have two prongs. The resultant neutrons formed in such a capture process have been observed by Sard *et al.* (1949).

4.3. *π-Mesons*

The slow positive π-meson is now known to decay into the μ-meson, but the slow negative π-meson is strongly nuclear reactive. Evidence for the interaction of fast π-mesons will be presented in §3.5, Part III (see also Powell 1950).

The results of the work of Perkins (1949) on stars formed by cosmic-ray π-mesons and of Gardner (1949) on artificially produced π-mesons are given in Table 3.

Table 3. Size Distribution of Stars produced by π-Mesons

No. of prongs n		0	1	2	3	4	5
% stars with n prongs	Perkins	29	30	19	14	7	1
	Gardner	27	23	24	15	9	2

It is seen that the results are closely parallel. The large percentage of stars in which no charged particles are emitted may be noted. By the use of sandwich emulsions made up of alternate layers of gelatine and emulsion and a study of the α-particle to proton ratios in stars produced in the two materials Perkins concluded that most of the stars with more than one prong are produced in the gelatine. He thought that an appreciable fraction of the one-prong stars might be produced in silver and bromine nuclei. Perkins' and Gardner's results have been extended by Menon, Muirhead and Rochat (1950), who have made a careful study of some 2,553 stars of which 1,337 came from artificially produced π-mesons and 1,216 from cosmic-ray π-mesons. They have separated the stars formed in light and heavy nuclei by the methods used by Perkins and also by noting that the ranges of the α-particles for light and heavy nuclei are different. The α-particles of energy less than 9 MeV. come from light materials and those above 9 MeV. from heavy materials. The results are shown in Table 4. It may be noted that 46% of the stars come from the light elements and 54% from the heavy elements.

Table 4. Percentage of π^--Stars of Different Types produced in Light Elements, L, and Heavy Elements, H

No. of prongs n		1	2	3	4	5	Total
% stars with n prongs	L	11	13	15	6	1	46
	H	27	18	4	4	1	54

The prong distributions are clearly very different in the two types. Light nuclei give a maximum at $n = 3$, whereas heavy nuclei show a gradual decrease as n increases. Menon, Muirhead and Rochat estimate that some 80% of the π^--mesons which stop in the emulsion without producing observable stars are due to capture by heavy nuclei. These authors have also measured the energies of many of the emitted particles and they find the surprising result that protons of energy greater than 30 MeV. are emitted in 4·9% of the stars observed in light nuclei and in 21% of the stars observed in heavy nuclei. For heavy nuclei particles below 30 MeV. have an energy distribution which may be fitted closely to a Weisskopf curve with average energy 9·6 MeV. per particle and mean excitation energy 109 ± 20 MeV. Thus evaporation is the main process by

which the rest energy of the captured π-meson is liberated. The size–frequency curve has also been computed on a number of models (e.g. Fujimoto, Hayakawa and Yamaguchi 1949, Clementel and Puppi 1949, Menon *et al.* 1950) and a good fit is found with the Fermi gas model and a suitable fluctuation formula.

The process of absorption in light nuclei seems to follow the main lines of absorption by heavy nuclei except that it is rather more complex. The actual identification of the light nuclei has been made in over half of all the stars examined. With ^{12}C and ^{14}N specific reactions are often observed. Alpha particles of energies greater than 20 MEV. are occasionally observed. No tritons of energies greater than 30 MEV. are observed.

These observations make possible a definite check of the theories of π^--meson capture. Perkins, for example, has assumed that the capture process involves a proton–neutron pair and leads to the formation of two neutrons each of energy 70 MEV. A mechanism of this type is needed to account for the high value of the excitation energy. Tamor (1950) has suggested capture by an α-particle group which later disintegrates with the formation of a 30 MEV. triton and a 90 MEV. neutron.

Tamor has calculated on these two models the percentages of protons with energies greater than 30 MEV. emitted per 100 mesons absorbed in light and heavy materials. On the proton–neutron model there should be 48 protons per 100 mesons for nitrogen and 24 for silver, and on the α-particle model 12 protons for nitrogen and 13 for silver. The experimental results definitely disagree with the first model but the average result agrees with the second model. There is, however, a definite disparity in the ratio of the results for light and heavy nuclei; further, no tritons of 30 MEV. are observed at all. It is suggested, following a model used by Ruddlesden and Clark (1949), that the process of absorption involves several nucleons which lead to the emission of a small fraction of fast nucleons and a general heating-up of the nucleus. A model is given for heavy nuclei which leads to approximately the correct percentage of fast nucleons and the correct value of the excitation energy. In light nuclei the primary process involves a large part of the nucleus and leads to the characteristic types which are observed.

III. THE TRANSITION EFFECTS OF THE NUCLEONIC AND NUCLEAR COMPONENTS

§ 1. INTERACTION MEAN FREE PATH AND ATTENUATION LENGTHS

It will be clear from the work on stars described in Part II that there is a generic connection between the nucleons in the atmosphere and the primary cosmic radiation. The primary cosmic-ray particle, which is a charged nucleon or a heavy particle, produces a cascade of penetrating showers and evaporation stars throughout the atmosphere, the extent of the penetration of the primary particle depending mainly on its energy. The final nucleonic component in the lower reaches of the atmosphere consists of slow protons and slow neutrons.

Fujimoto and Hayakawa (1949) have pointed out that it is possible to distinguish three nucleonic components, namely, the primaries of penetrating showers (component A), the primaries of evaporation stars (component B), and the final nucleonic products of evaporation stars (component C). It will be convenient to use this classification. If it is necessary to refer to the whole nucleonic component, irrespective of its type, the term N-rays, suggested by Rossi, will be used.

We now wish to examine the interaction of these various nucleonic components with matter and discuss their connection with one another in the earth's atmosphere and in condensed matter. In this connection it is important to define interaction mean free path and attenuation * lengths.

The *interaction mean free path* (which is sometimes termed the collision mean free path), denoted by λ_I, is defined as the average distance an N-ray goes before it produces its first nuclear interaction.

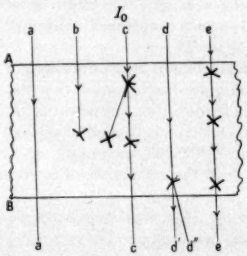


Figure 17. Schematic representation of the attenuation of N-rays by matter.

The *attenuation length*, denoted by λ_A, is defined as the thickness of absorber which reduces the intensity of N-rays by a factor $1/e$. Thus if I_0 is the intensity of N-rays at A (Figure 17) and I the intensity at B, x gm/cm^2 below A, the relation between I and I_0 is given by

$$I = I_0 \exp(-x/\lambda_A). \qquad \ldots\ldots(1.1)$$

It is found experimentally that the attenuation of N-rays is exponential, as assumed in the equation. The interaction mean free path is defined from the relation

$$I_I = I_0 \exp(-x/\lambda_I), \qquad \ldots\ldots(1.2)$$

where I_I is now the intensity of emergent rays of the type a. Rays c, d and e are excluded because each of these has reached the level B via one or more nuclear interactions.

In practice, λ_A is always greater than λ_I. When the absorption process is catastrophic, $\lambda_A = \lambda_I$.

It is important to realize that the incident rays are not all of the same energy. The attenuation length refers therefore to a spectrum of N-rays which may change as the rays pass through the absorber. Moreover, the interaction mean free path may be a function of energy and of the type of interaction undergone by the rays.

The interaction mean free path corresponding to the geometrical cross section of the target nucleus is given by equation (1.3),

$$\lambda_I = 1/\Sigma n_j \sigma_j \qquad \ldots\ldots(1.3)$$

where n_j is the number of nuclei of kind j per unit mass of the absorber and σ_j is the cross section of this nucleus. The geometrical cross section of a nucleus is defined by equation (1.4):

$$\sigma = \pi r_0^2 A^{2/3}, \qquad \ldots\ldots(1.4)$$

* Following Mylroi and Wilson (1951), the term 'attenuation' is used in preference to 'absorption'.

where r_0 is the range of nuclear forces of the proton, and is equal to $1{\cdot}45 \times 10^{-13}$ cm., and A is the atomic weight.

Equation (1.3) assumes that the incident particle is a point. For a primary proton this assumption is equivalent to neglecting the range r_0 of nuclear forces of the proton.

§ 2. THE INTERACTION MEAN FREE PATHS OF THE HEAVY PRIMARIES

The interaction mean free paths of the heavy cosmic-ray primaries have been determined by Bradt and Peters (1950). An interaction is defined as a reduction in the charge of the nucleus by two units. The results for three groups of elements are shown in Table 5. Group I includes carbon, nitrogen and oxygen, Group II elements like silicon and magnesium, and Group III iron. The method is simply to count the number of collisions occurring in a given length of path through glass. These quantities are given in columns (3) and (4) of the Table.

Since the heavy nucleus has a finite size, the geometrical cross section is defined by

$$\sigma_j = \pi(R_j + R_i)^2 \qquad \ldots\ldots(2.1)$$

where R_i is the radius of the incident nucleus and R_j is the radius of the target nucleus. The interaction mean free path corresponding to the geometrical cross section is then defined by equation (1.3). The calculated values of λ_I are given in column (6) of Table 5.

Table 5. The Interaction Mean Free Paths of Heavy Primaries

(1)	(2)	(3)	(4)	(5)	(6)	(7)
I	6–8	40	520	33 ± 5	20	34
II	10–18	26	255	25 ± 5	16	25
III	26	8	59	19 ± 6	12	18

(1) Group; (2) Z; (3) No. of collisions; (4) Total path length (cm.); (5) λ_I (gm.cm^{-2}), observed; (6) λ_I (gm.cm^{-2}), $\Delta R = 0$ (eqn. (2.1)); (7) λ_I (gm.cm^{-2}), $\Delta R = 0{\cdot}85 \times 10^{-13}$ (eqn. (2.2)).

It will be noted that the observed values of λ_I are much larger than the calculated. Bradt and Peters therefore suggest that there might be some overlapping of the geometrical radii without nuclear interaction. If the correction for each nucleus is ΔR, the effective radii are $(R_j - \Delta R)$ and $(R_i - \Delta R)$ respectively. Thus the effective cross section is given by

$$\sigma_{\text{eff}} = \pi(R_j + R_i - 2\Delta R)^2. \qquad \ldots\ldots(2.2)$$

With $\Delta R = 0{\cdot}85 \times 10^{-13}$ cm. the calculated and observed values agree well (see column (7)). This decrement is approximately equivalent to neglecting the cross section of the incident nucleon in equation (1.3).

The values of the interaction mean free path are all very short and, as will be seen later, are much shorter than for the primary protons. Thus heavy nuclei are absorbed very rapidly in the earth's atmosphere.

§ 3. THE INTERACTION MEAN FREE PATHS OF THE PRIMARIES AND SECONDARIES OF PENETRATING SHOWERS

3.1. *Introduction*

In § 2 it has been shown that the mean free paths of the heavy particles can be measured for any particular type of interaction by counting the number of collisions in a given length of track. This method can be used because a unique determination

can be made of the type of heavy primary at all points along its track. The method cannot be applied to particles of charge e above energies of about 1,000 MeV. because all these particles ionize at approximately the minimum value and no discrimination is possible. Discrimination is possible below 1,000 MeV., and in § 3.5 the preliminary results of the application of the photographic plate to this problem will be discussed. Most primary particles (class A) lie in the energy region above 1,000 MeV. The main evidence for the interaction lengths of these particles comes from counter and cloud-chamber experiments on penetrating showers.

3.2. *The Transition Effect of Penetrating Showers*

It would seem that in principle the interaction length of very energetic primaries could be measured by studying the transition effect of penetrating showers. This could be carried out with a counter set of the type shown in Figure 9 by observing the rate of local penetrating showers as a function of the thickness of the transition material T. If it is assumed that the secondary shower particles have greater range than the dimensions of the apparatus (and the range of the primaries), and that one or more showers produced by the *same* primary in the layer T are recorded as one event, it is easily shown that the rate of penetrating showers is proportional to $[1-\exp(-x/\lambda_1)]$.

Transition curves of this general form are in fact found (Jánossy and Rochester 1944 a), and they lead to a value of of 60–100 gm. cm^{-2} for lead. It is clear, however, from cloud-chamber photographs that large penetrating showers are often complex phenomena and, therefore, the true value of λ_1 may be very different from the one calculated from the simple inverse exponential form. George and Jason (1950) have examined some of the factors involved. They show that the main correction to be applied comes from the difference in efficiency of the shower set for recording showers produced in different parts of the transition material. With this correction, which they determine experimentally by placing the transition layer T at different distances above the shower set, they find λ_1 for lead to be 180 ± 40 gm. cm^{-2} and for aluminium 85 ± 15 gm. cm^{-2}. These values are consistent with mean free paths corresponding to cross sections equal to the geometrical cross sections. Since a shower set of this type is set off only by very high energy nucleons, perhaps as much as 80,000 MeV. (Walsh and Piccioni 1950), the measured cross sections, refer only to particles of this type. The results should, however still be taken with reserve because of the other factors, such as the different recording efficiencies of the various counter trays and the effects of secondary reactions throughout the lead, for which allowance cannot be made.

3.3. *The Determination of the Interaction Mean Free Path of Charged Primaries*

In experiments of the type considered in § 3.2 the transition layer acts both as an absorber of the primary particles and as a source of the showers which set off the sevenfold set. Cocconi (1949 a, b) devised a counter arrangement, shown in principle in Figure 18, which will separate these two effects and thus permit a more accurate determination of λ_1. The counters are so arranged that only shower events in P, for which once ounter in tray A is discharged, are recorded. Events of type b are therefore excluded. This arrangement is better than the sevenfold shower set, but even so the results are complicated by the recording of spurious events, like those shown in Figure 19, which simulate single charged primaries. Cocconi considers that for lead these spurious events only lead to a small background.

There is also an important effect due to knock-on showers from μ-mesons, but this effect can be measured and allowed for.

The actual arrangement used by Cocconi at altitudes of 3,260 m. and 260 m. is shown in Figure 20. Counters B, C and D, which record the penetrating showers, are each separated by 1 inch of lead. Coincidences of the type $A+B+C+D+1E$ are taken as a function of the thickness of absorber $\Sigma'+\Sigma$. Thus to produce

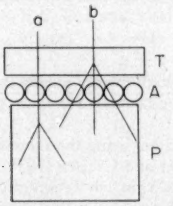


Figure 18. Diagram to illustrate the counter method of determining the interaction mean free path of the primaries of penetrating particles. The counter set P records penetrating showers. The counters A are hodoscoped so that events of type a can be recorded and those of type b excluded. (*After* Rossi 1949)

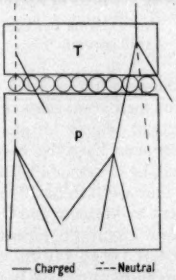


Figure 19. Diagram to illustrate the spurious events which may be recorded by the counter method of measuring the interaction mean free path.

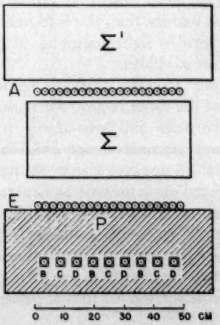


Figure 20. Arrangement of counters and absorbers used by Cocconi to measure the interaction mean free path. The penetrating shower detector consists of counter sets BCD embedded in a large mass of lead. Only those penetrating showers which are accompanied by the discharge of one counter in tray A and one in tray E are recorded. (*After* Cocconi 1949 a)

a shower a ray has to set off only one counter in tray A and only one in tray E. After correcting for the knock-on shower background Cocconi finds $\lambda_I(\mathrm{Pb}) = 160 \pm 15$ gm. cm^{-2}.

Similar experiments have been carried out by Tinlot and Gregory (1949), Walker (1950) and Sitte (1950 a). Walker and Sitte have extended the Cocconi work by determining the mean free path for the charged primaries which produce

showers with different numbers of relativistic secondary particles. Both workers find a decrease in the value of λ_I with increasing shower size. Walker's results are shown in Table 6. Sitte finds $\lambda_I(\text{Pb}) = (196 \pm 13)$ gm. cm^{-2} for secondaries which can penetrate 100 gm. cm^{-2} of absorber and $\lambda_I(\text{Pb}) = (162 \pm 10)$ gm. cm^{-2} for secondaries which can penetrate 200 gm. cm^{-2} of absorber.

Table 6. Values of λ_I(Pb) as a Function of Number of Shower Particles

No. of shower particles	3	4, 5	>5	>7
λ_I(Pb) (gm. cm^{-2})	208	180±10	150±8	140±10

All the experiments seem to indicate that for high energy events in lead the interaction mean free path is consistent with the value calculated from the geometric cross section, namely 156 gm. cm^{-2}.

Experiments have also been carried out for carbon, using the same apparatus as for lead, by Cocconi (1949 b) and Walker, Walker and Greisen (1950). These workers find λ_I(C) to be 100 ± 5 gm. cm^{-2} and 81 ± 5 gm. cm^{-2} respectively. An experiment by Mezzetti and Querzoli (1950), which requires at least two particles to penetrate 7·5 cm. Pb, gives the ratio $\lambda_I(\text{Pb}) : \lambda_I(\text{C})$ as $4{\cdot}4 \pm 1{\cdot}0$, which leads to a value for λ_I(C) of 48 ± 10 gm. cm^{-2}, since the value for λ_I(Pb) for the conditions of their experiment must have been about 210 gm. cm^{-2}. The value of λ_I(C) corresponding to the geometrical cross section is about 60 gm. cm^{-2}. Thus the value found by the Italian workers is consistent with the length corresponding to the geometrical cross section, while the American result indicates a significantly smaller cross section. The latter result, if correct, probably means that the carbon nucleus is more transparent than the lead nucleus. Transparency for the scattering of 90 Mev. neutrons by light nuclei has already been found at Berkeley (see Bethe 1940, Fernbach *et al.* 1949).

Davies, Franzinetti and Perkins (private communication) have recently investigated the problem of the cross section for meson production by measuring the energies of shower primaries and secondaries in stars in the photographic emulsion. Of 155 primary particles, each with a total energy greater than $1{\cdot}69 \times 10^3$ Mev., 65 particles did not give rise to shower particles. Some of these particles might have produced slow mesons or neutral mesons, but these effects are likely to be small. Their results are shown in Figure 21, where the probability of producing a shower particle (inelastic collision) is plotted as a function of the kinetic energy of the primary particle. It is seen that the probability of producing mesons is large only at energies greater than 5×10^3 Mev. The curve for inelastic collisions shows how the cross section for meson production varies with energy.

3.4. *The Determination of the Interaction Mean Free Path of Neutral Primaries*

The determination of λ_I for neutral primary particles may be made with an apparatus of the Cocconi type except that the counters of tray A have now to be connected in anticoincidence with the shower set P.

The apparatus used by Rossi and Regener (1940) is illustrated in Figure 22. Counters B, C, D and E form a counter telescope and tray A is connected in anticoincidence with these counters. The absorber R consists of 5 cm. of lead and it can be placed in positions R_1 or R_2. A coincidence $B+C+D+E$, unaccompanied by a discharge in A, is designated $B+C+D+E-A$. The experiment

consists in observing the change in the rate of this type of coincidence with the thickness of the absorber Σ when 5 cm. of lead are in the position R_1 and no lead is in position R_2. It is assumed that these counts are due mainly to the production of charged penetrating particles in R_1 by neutral radiation passing through the counter bank A. The counts observed with 5 cm. of lead in position R_2, and no absorber in R_1, are assumed to give the background counts, which are due to the leakage of charged particles through counters A, side showers and penetrating particles scattered into B, C, D, E and missing A. A large increase in the count B + C + D + E − A is found when the 5 cm. lead is moved from position R_2 to position R_1, approximately one-third of the counts being accompanied by the simultaneous discharge of counters F. Assuming a constant background count for the two positions R_1 and R_2, Rossi (1949) deduced a value of λ_1 for lead of 130 gm. cm^{-2}. In the nomenclature of Brown *et al.* (1949) the type of event observed is clearly 1n + 2n + 3n etc.

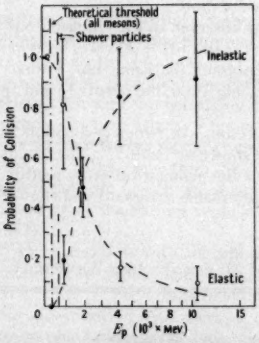


Figure 21. The probability of producing a shower particle in a nuclear interaction as a function of the energy E_p of the primary in MeV. × 10^3. (*After* Davies, Franzinetti and Perkins, private communication)

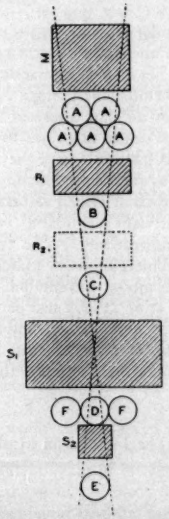


Figure 22. The counter arrangement used by Rossi and Regener for the investigation of the production of penetrating particles by non-ionizing rays. Counters B, C, D and E form a counter telescope and the tray A is connected in anticoincidence with these counters. Counters F are used for recording showers. The experiment consists in determining the rate of anticoincidences B + C + D + E − A as a function of the thickness of absorber Σ when $R_1 = 5$ cm. lead and $R_2 = 0$. These counts are due mainly to neutral radiation passing through A and producing ionizing penetrating radiation in R_1. (*After* Rossi and Regener 1940)

The value of λ_I for neutral penetrating shower primaries has been determined by Walker *et al.* (1950). For lead the value is 164 ± 15 gm. cm^{-2} and for carbon 80 ± 7 gm. cm^{-2}. These results are not significantly different from the values for charged primaries.

Walker *et al.* (1950) also find the ratio of neutral to charged primaries of penetrating showers to be $0{\cdot}84 \pm 0{\cdot}13$.

3.5. *The Secondary Interactions of the Particles produced in Stars in the Photographic Emulsion*

The secondary interactions of some of the particles produced in stars have been observed by Camerini, Fowler, Lock and Muirhead (1950) using electron-sensitive emulsions. Six nuclear interactions have been found in 128 cm. of π-meson track length in the emulsion, giving a value of $\lambda_I = 82 \pm 35$ gm. cm^{-2} of emulsion. The value of λ_I for *all* thin tracks of length greater than $400\,\mu$ is 102 ± 27 gm. cm^{-2}. Both of these values are close to the value expected from the geometrical cross section, i.e. 90 gm. cm^{-2}, and they lead to the important conclusion that the fast π-meson has as great an interaction with matter as the proton.

Out of a total of fifteen secondary interactions observed in the emulsion only four have relativistic secondaries, two being of the type 1p and the others of the types 3p and 5p respectively. The type 1p may represent the same fast primary entering and emerging from a nucleus after producing a star, but events 3p and 5p must indicate meson production.

Camerini *et al.* (1950) also determine λ_I for protons, deuterons and tritons in certain specific energy ranges, and their results are shown in Table 7. The results are again consistent with a cross section equal to the geometrical cross section. Also included in the Table are the results for identifiable π-mesons and shower particles.

Table 7. The Interaction Mean Free Paths of Mesons, Protons, Deuterons and Tritons

Particle type	Energy range (Mev.)	Path length (cm.)	λ_I (cm. of emulsion)
Proton	40–100	174	32 ± 16
Proton	100–600	291	30 ± 10
Deuteron + triton	75–250	127	$12 \pm 4{\cdot}5$
Mesons	150–1500	128	21 ± 9
Shower particles:			
(Mesons)	150–5000	389	26 ± 7
(Protons)	600–5000		

Camerini *et al.* have informed the authors in a private communication that they have now 40 stars caused by identified protons and 40 by identified π-mesons. Of the stars produced by protons, eleven are in the energy range 600–850 Mev., and in only one case is an identified π-meson produced. From the consideration of the energy balance there appears to be no evidence for 'missing' energy. Of the stars produced by π-mesons, 22 are caused by mesons with energies of from 600 to 1,100 Mev. In six cases one π-meson also comes out of the star, and in five of these the energy of the emitted meson is about 4% of the incident π-meson. In the remaining case the emergent π-meson has half the energy of the incident meson.

Considering the energy balance of all of the 40 stars, there is evidence of 'missing' energy. It is therefore possible that neutral π-mesons are produced as well as charged mesons.

3.6. *The Large-Angle Scatterings and the Secondary Interactions of the Penetrating Particles observed in the Cloud Chamber*

The interaction mean free path may also be determined by using a cloud chamber containing one or more metal plates and triggering it by a counter arrangement which selects penetrating showers. If N_0 is the number of plates of thickness l gm. cm^{-2} penetrated by the shower particles and N is the number of large-angle scattering and nuclear interactions observed, λ_I is given by

$$\lambda_I = (N_0 l/N) \text{ gm. cm}^{-2}. \qquad \ldots\ldots (3.61)$$

A large-angle scattering is defined by Barker and Butler (1951) as one which is more than five times the expected Williams' scattering and by Fretter as an angle greater than 15°. These criteria are somewhat arbitrary, especially when the momentum of the scattered particle cannot be measured directly. It is assumed that a large-angle scattering represents a nuclear interaction, and that if evaporation particles are produced they are not seen because of absorption in the metal plates. The use of thick metal plates also leads to a loss of nuclear interactions when the interactions lead only to the production of evaporation particles and no shower particles. Brown and McKay (1950) have given a method for correcting for this loss. Nevertheless, even with this correction it is probable that the values of λ_I represent upper limits to the true values. Some recent values for lead are included in Table 8, from which it is seen that after correction for loss of stars the observed interaction mean free paths range from 200–320 gm. cm^{-2}. The particles observed in these experiments are usually an unknown mixture of π-mesons and protons; a partial separation of the various types of particle has been made by Barker and Butler (1951).

Here again, as with the photographic plate work, the observed interactions include evaporation stars, meson production and large-angle scatterings. No attempt has yet been made to obtain λ_I for the different types of interaction and to study its variation with energy. It is clear that the observed mean free path for meson production, at these energies, is appreciably greater than the value expected from the geometrical cross section.

Table 8. Values of λ_I(Pb) from Observations in Cloud Chambers

Authors	Altitude (m.)	λ_I (gm. cm^{-2})
Fretter (1949 a)	3027	750 (uncorrected)
Lovati *et al.* (1950)	3450	300 ± 100 (corrected)
Brown and McKay (1950)	3260	316 ± 70 (corrected)
Butler *et al.* (1950)	Sea level	400 ± 150 (uncorrected)
Barker and Butler (1951)	Sea level	200 ± 80 (corrected)

§ 4. THE ATTENUATION OF HIGH-ENERGY NUCLEONS IN THE ATMOSPHERE

4.1. *Introduction*

In this section we shall be concerned with experiments on penetrating showers in the atmosphere which lead to a determination of the attenuation length of high-energy nucleons. The reason why the attenuation length and not the interaction mean free path is obtained will be clear from Figure 23, in which is shown a primary

cosmic-ray initiating a series of nucleon cascades through the atmosphere. From the evidence presented in §3 it is clear that the interaction mean free paths of all the fast nucleons in the cascade must be at least as great as the value corresponding to the geometrical cross section; the result is that in air there will be large distances between interactions (at sea level the value of λ_I (air) must be of the order of hundreds of metres of air). Thus even if the fast nucleons are emitted at the top of the atmosphere with quite a small angular spread, the spatial separation of the rays in the lower reaches of the atmosphere must be much greater than the dimensions of any recording unit, whether it is a penetrating shower set or a photographic plate. The rate of a penetrating shower recorder in the atmosphere will therefore be proportional to the total flux of fast N-rays at the point of observation, and the determination of the rate of penetrating showers as a function of altitude will give λ_A. If the absorption is exponential, equation (1.1) of Part III may be used, but at very high altitudes, where the fast N-rays are isotropic, the variation in shower primaries with depth is given by the Gross transformation,

$$I = I_0\left\{\exp(-x/\lambda_A) - \frac{x}{\lambda_A}\int_{x/\lambda_A}^{\infty}\frac{e^{-z}}{z}\,dz\right\}, \qquad \ldots\ldots(4.11)$$

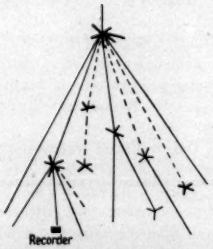


Figure 23. Schematic representation of nucleon cascades in the atmosphere. The numbers of heavily ionizing particles coming from two of the stars are exaggerated.

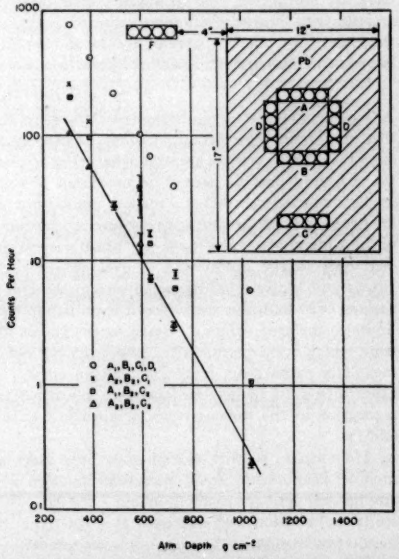


Figure 24. Counter arrangement (inset) and results of Tinlot on the variation of penetrating showers with altitude. Shower events are classified by symbols of the type $A\alpha$, $B\beta$ . . . . where α and β represent the numbers of counters discharged in trays A, B . . . . (*After* Tinlot 1948 a)

where x is the depth below the top of the atmosphere. This equation approximates to exponential form in the lower atmosphere.

Finally, it may be noted that since the primary spectrum is cut off in northern latitudes at an energy of about 2,000 MEV., all the fast nucleons which have energies lower than this value must be of secondary origin. Again, since the primary rays are charged particles, most of the fast neutral nucleons must be secondaries.

4.2. *Counter Experiments on the Variation of Penetrating Showers with Altitude*

The arrangement used by Tinlot (1948 a, b) is shown in the inset to Figure 24. In the experiment the rates of variation with altitude of coincidences of the type $A_\alpha B_\beta C_\gamma D_\delta$ are first investigated, and it is found that events other than penetrating showers are effectively eliminated with a sixfold set of the type $A_2 B_2 C_2$ or a fourfold grouping $A_1 B_1 C_1 D_1$. Extensive showers can be excluded by the use of a tray of unshielded counters at F. Tinlot finds that the rate of local penetrating showers increases exponentially with height up to the maximum altitude (9,000 m.) reached by his aeroplane, giving $\lambda_A(\text{air}) = 118 \pm 2 \text{ gm. cm}^{-2}$. Results consistent with this value have been obtained by Sala and Wataghin (1946), George and Jasôn (1948) and Hodson (private communication). The rates of penetrating showers at altitudes of 33,000 ft. and 11,000 ft. are about 650 and 25 times the rates at sea level.

4.3. *Photographic-Plate Experiments on the Variation of High-Energy Stars with Altitude*

The most extensive data yet obtained on the variation of high-energy stars with altitude seem to be those of Camerini *et al.* (1949) and Lord and Schein (1950). Both groups of workers have sent electron-sensitive plates up in balloons, in one case to an altitude of 70,000 ft. and in the other to 95,000 ft. (15 gm/cm² residual atmosphere). The results of Camerini and his co-workers on the variation of the number of stars per cm³ per day as a function of the number of shower particles n_s are shown in Figure 25. On the same diagram are plotted the results of Brown *et al.* (1949) obtained at 11,000 ft. and, for comparison, the results at 70,000 ft. normalized to correspond to a total rate of production equal to that at 11,000 ft. (dotted curve). Not only do these curves show a great increase in the star rate with altitude but also a striking change in the distribution of star types. There is clearly a marked increase in the number of high-energy events at the higher altitude. This change reflects a difference in the spectra of high-energy nucleons at the two altitudes and must be due mainly to the fact that the cut-off imposed by the earth's magnetic field allows only nucleons of energy greater than 2,000 MEV. to enter the top of the atmosphere (in northern latitudes). Thus there are relatively fewer showers of the types $n_s \leqslant 3$ at 70,000 ft. than at 11,000 ft. because the nucleons which produce these types are secondary particles. Another fact which is consistent with this picture is the preponderance of penetrating showers produced by charged nucleons at high altitudes. Lord and Schein find at 95,000 ft. that 90% of the stars which have tracks at minimum ionization are produced by charged particles; at 11,000 ft. approximately equal numbers are produced by charged and uncharged particles (Brown *et al.* 1949, and § 2.1, Part II). Again, the angular distribution of the primaries of the high-energy stars becomes more nearly isotropic the higher the altitude. This change in angular distribution makes it necessary to use the Gross

transformation to determine λ_A. The values found by Camerini *et al.* for different star sizes are shown in Table 9. It will be observed that the decrease in λ_T with the energy of the star, noted in § 3.3, is paralleled here for λ_A. The values of λ_A for the largest stars are consistent with Tinlot's value of 118 gm. cm^{-2}.

Table 9. Values of λ_A for Stars of Different Type

Star type n_s	0	3	6
Value of λ_A (gm.cm^{-2}):			
Gross transformation (eqn. (4.11))	170	135	120
Exponential variation (eqn. (1.1), Part III)	130	105	95

Finally, the frequency–size distributions of stars at 95,000 ft. as determined by Lord and Schein are shown in Figure 26. Curve A gives the distribution for particles of all types and curve B for shower particles only. Marked changes in slope occur at $n=6$. Comparison of curve A with Figure 1 emphasizes again the relative increase in the high-energy stars at the higher altitudes.

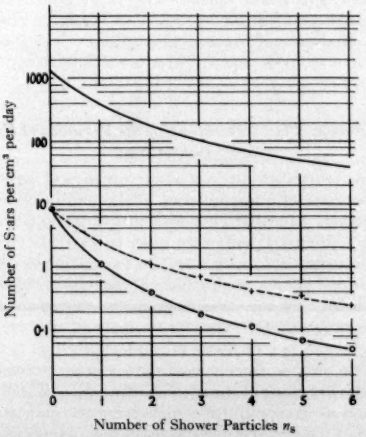


Figure 25. Rates of production of stars with different values of n_s at 70,000 ft. and at 11,000 ft. The broken line shows the results for 70,000 ft. normalized to correspond to a total rate of production equal to that observed at 11,000 ft. (*After* Camerini *et al.* 1949)

4.4. *The Barometer Effect of Penetrating Showers*

A marked barometer effect of penetrating showers at sea level has been found by Jánossy and Rochester (1944 b) using a sevenfold penetrating shower set of the type illustrated in Figure 9. The barometer coefficient is equal to $-11{\cdot}7 \pm 2{\cdot}7\%$ per cm. Hg. A change in barometric pressure means a change in weight of air above the apparatus and is equivalent to a change in thickness of the atmosphere at constant pressure. Thus the quantity λ_A is again determined. A value of $\lambda_A = 113 \pm 3$ gm. cm^{-2} is found, which is quite consistent with other determinations (Jánossy 1948).

4.5. *The Angular Distribution of the Primaries of Penetrating Showers*

Greisen (1942) has shown that if the zenith angle distribution of primary particles is proportional to $\cos^n\theta$, the angular distribution projected on to a vertical plane is proportional to $\cos^n\phi$ where ϕ is the projected zenith angle. Thus the

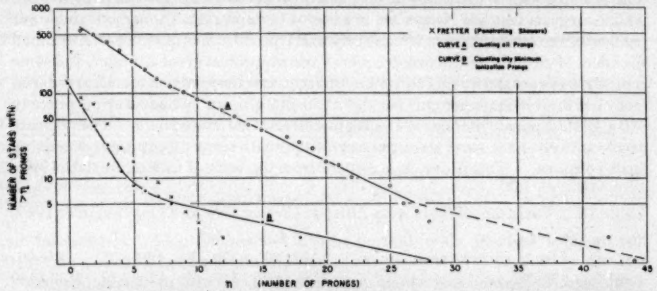


Figure 26. Integral size distribution of stars obtained in electron sensitive plates at an altitude of 95,000 ft. Curve A gives the distribution for all prongs and curve B for stars with more than n relativistic tracks. (*After* Lord and Schein 1950)

intensity of primaries at a depth l in the atmosphere in any direction is proportional to $\exp(-l/\lambda_A\cos\phi)$. For $n=l/\lambda_A\leqslant 8$, $\phi=30°$, $\exp(-l/\lambda_A\cos\phi)$ is approximately equal to $\exp(-n)\cos^n\phi$. This expression has been shown by Walker (1950) to fit the distribution of the primaries of penetrating showers at an altitude of 3,250 m. The value of n is $7\pm 1{\cdot}2$, giving $\lambda_A=100\pm 17$ gm. cm^{-2}.

§ 5. THE ATTENUATION OF STAR-PRODUCING NUCLEONS IN THE ATMOSPHERE AND IN CONDENSED MATTER

5.1. *Introduction*

In the previous sections we have been concerned mainly with the interaction and attenuation of the A component. We wish now to examine the experimental evidence relating mainly to the attenuation of the B component in the atmosphere and in condensed matter. The average energy of the B-type nucleons is about 200 MeV. Since at this energy the range of a proton in air is 20 gm. cm^{-2}, a proton is usually absorbed by ionization loss before it can produce a star. Consequently it is assumed that most of the B-rays consist of neutrons.

The correlation of the various components will be considered in Part IV.

5.2. *Photographic-Plate Experiments on the Variation of Evaporation Stars with Altitude*

Experiments of this type involve the determination of the rate of stars per cm^3 of emulsion per day as a function of altitude. The rate measured at any given altitude is proportional to the total flux of the A and B components. The variation with altitude clearly gives the attenuation length since there is no discrimination between particles which have reached the photographic plate via one or more nuclear interactions.

Bernardini, Cortini and Manfredini (1949) investigated the variation in stars produced in Ilford C2 plates up to an altitude of 20,000 m. and they obtained a value of the attenuation length of 135 ± 4 gm. cm^{-2}. George and Jason (1949) using similar plates at an altitude of 3,457 m. found $\lambda_A = 150 \pm 7$ gm. cm^{-2}. The size–frequency distribution changes little between sea level and 20,000 m., which suggests that the N-rays are in a state of equilibrium throughout the lower reaches of the atmosphere. Freier, Ney and Oppenheimer (1949) have examined the effect of altitude on the size–frequency distribution at great altitudes, and some typical results are shown in Table 10. It is seen that the total rate of stars is reduced from $1,870 \pm 140$ stars per cm^3 per day at 50–300 gm. cm^{-2} residual atmosphere to 880 ± 90 stars per cm^3 per day at 14–24 gm/cm^2 residual atmosphere. The decrease occurs entirely in the small stars, the ratio of stars with six or more prongs remaining fairly constant. This is due, as is evident from the work of Camerini *et al.* (1949),

Table 10. Variation of Stars with Altitude (Freier, Ney and Oppenheimer 1949)

Star type (no. of prongs)		3–5	6–9	⩾10	⩾3
Star rate (cm^3/day)	14–24 gm.cm^{-2} resid. atmos.	330±60	330±30	250±25	880±90
	50–300 ,, ,, ,,	1260±120	350±60	260±50	1870±140

to the fact that the small stars (of 3–5 prongs) are due mainly to the B component, and since this component is of secondary origin it passes through a maximum near the top of the atmosphere. The small change in the number of large stars with atmospheric depth near the top of the atmosphere is due to the fact that these stars are produced by a mixture of primary cosmic-rays and fast secondary nucleons. The decrease in the number of primary particles is compensated for by the generation of fast secondary nucleons capable of producing quite large stars.

5.3. *Ionization-Chamber Experiments on the Variation of Low-Energy Stars with Altitude*

The various events recorded by an ionization chamber were outlined in Part II, §2.3. With an *unshielded* chamber the recorded bursts at high altitudes are due almost entirely to nuclear interactions in the walls of the ionization chamber. Bridge, Hazen, Rossi and Williams (1948) find that the rate of rise in the atmosphere up to 10,000 m. gives a value of λ_A of 138 gm. cm^{-2}. The use of the Gross transformation increases this value to 165 gm. cm^{-2}.

The distribution of burst size with altitude, as measured by Coor and Snow (1948) and Hulsizer (1948), is in agreement with the results from photographic emulsion experiments.

5.4. *Transition Effects in Condensed Matter*

Transition effects in condensed matter have been investigated by many authors using both the photographic plate and the ionization chamber technique. Here again the attenuation length is determined. Some determinations of λ_A by the photographic plates method are given in Table 11.

Bernardini, Cortini and Manfredini (1949) have observed for small stars a transition effect which has a small maximum at a thickness of 2 cm. of lead. Malaspina *et al.* (1950) find a marked increase in the number of small stars recorded by a plate when the absorber was placed beneath the plates. This suggests that the

transition effect is closely correlated to the N-rays, which are almost isotropic in the absorber. The transition effect has been interpreted by Dallaporta *et al.* (1950) as due to the secondary neutrons produced in evaporation stars in the absorber. Barton, George and Jason (1950) have recently found that the attenuation length in air is shorter than in carbon, i.e. 143 ± 5 gm. cm^{-2} as compared with 166 ± 7 gm. cm^{-2}, although air is a substance of higher atomic weight. The difference is attributed to the stars produced by π-mesons created locally in the

Table 11. The Attenuation Length (gm.cm^{-2}) in Different Materials

Authors	Material			
	Carbon	Ice	Aluminium	Lead
Barton, George and Jason (1950)	166 ± 7			
George and Jason (1949)				310 ± 20
Bernardini *et al.* (1949)			220	300 ± 20
Harding, Lattimore, Li and Perkins (1949)		200 ± 10		

carbon absorber. Such mesons will also be formed in air, but they will usually decay into μ-mesons before producing stars.

Determinations of λ_A for lead and iron using the ionization chamber method have been made by Bridge and Rossi (1949) with apparatus already described (see Figure 16). These authors find $\lambda_A(\text{Pb}) = 430 \pm 90$ gm. cm^{-2} and $\lambda_A(\text{Fe}) = 320 \pm 70$ gm. cm^{-2}. Bursts from μ-mesons may contribute somewhat to the measured values.

§6. THE LATITUDE EFFECTS OF PENETRATING SHOWERS AND STARS

6.1. *Penetrating Showers*

The latitude effect of penetrating showers has been investigated by Appapillai and Mailvaganam (1948, 1950) at sea level and by Walsh and Piccioni (1950) at altitudes of 25,000 ft. (383 gm. cm^{-2}) and 30,500 ft. (300 gm. cm^{-2}). Both groups of workers use counter arrangements sensitive to penetrating showers: in the first case a typical sevenfold set (see Figure 9) and in the second a penetrating shower set of the Tinlot type. Appapillai and Mailvaganam find a latitude effect of $0 \pm 15\%$ between geomagnetic latitudes 56·5° N. and 4° S., and Walsh and Piccioni (at both altitudes) an effect of $(12 \pm 2)\%$ between geomagnetic latitudes 55·13° N. and 20·7° N. Between the latter latitudes the ratio of the μ-meson components is 1·24 at 24,000 ft. and 1·33 at 30,500 ft. The absence of an appreciable latitude effect indicates that the primaries of penetrating showers are themselves of very high energy, i.e. greater than 10^{10} ev., or else that they come directly from primaries of very high energy. Walsh and Piccioni deduce from the results that a sevenfold penetrating shower set is set off by particles of energies about 8×10^{10} ev., and that their apparatus requires nucleons of energies greater than 2×10^{10} ev.

6.2. *Stars*

The latitude effect for stars has been measured by Yagoda (1949), Salant (1949), Simpson (1949) and Lattimore (1950). Yagoda's results are shown in Table 12(*a*), from which it will be observed that the rate of stars, at an altitude of 6,267 ft., falls by a factor of 2·6 between geomagnetic latitudes 55° N. and 39° N. All workers find an increase in the proportion of large stars near the equator and a

gradual decrease in the latitude effect with increasing size of star. Some observations of Lattimore with Ilford G5 plates exposed at an altitude of 11,500 ft. given in Table 12 (*b*) show the ratio of star rates at the Jungfraujoch and the Andes. The decrease in the latitude effect for larger stars is consistent with the results obtained for penetrating showers and may be explained in the same way. The magnitude of the latitude effect for the smaller stars means that an appreciable fraction is produced by nucleons which are the progeny of primary cosmic-ray particles between the cut-off energies corresponding to the geomagnetic latitudes at which the observations have been taken. Lattimore has pointed out that his observations require about half of the stars to have been produced by primary nucleons with energies between 3,000 MEV. and 15,000 MEV. This indicates that the other half of the stars are produced by primary nucleons of energies greater than 15,000 MEV. Since, however, there are probably about fifty times as many primary particles in the energy range 3,000–15,000 MEV. as above 15,000 MEV., nucleons with energies greater than this value must produce large numbers of neutrons in the atmosphere.

Table 12

(*a*) The Latitude Effect of Stars (Yagoda 1949)

Geomagnetic latitude (° N.)	55	42	39
Star rate (cm³/day) at 6,267 ft.	5·8	3·8	2·2

(*b*) The Decrease in the Latitude Effect with Increasing Star Size (Lattimore 1950)

Star size (no. of prongs)	3, 4	5, 6	7; 8	>8	All stars
Star rate (Jung.)/star rate (Andes)	$2{\cdot}6\pm0{\cdot}3$	$1{\cdot}9\pm0{\cdot}3$	$1{\cdot}2\pm0{\cdot}3$	$1{\cdot}1\pm0{\cdot}3$	$2{\cdot}1\pm0{\cdot}4$

IV. THE NUCLEONIC COMPONENTS IN THE ATMOSPHERE AND THEIR CORRELATION

§ 1. INTRODUCTION

As outlined in § 4, Part III, the absorption of a high-energy cosmic-ray primary particle in the atmosphere involves the production of a cascade of penetrating showers and evaporation stars, the final nucleonic products being mainly slow protons and neutrons. Direct evidence for this general picture comes from photographs of penetrating showers taken with cloud chambers containing many lead plates. Two recent examples of such cascades by Fretter and Lovati and his co-workers are reproduced in Plates VII and VIII. From such pictures it is clear that the difference between the attenuation length and the interaction length is due to successive production of stars by the incident nucleon and its fast progeny. In general, the attenuation length is approximately twice the interaction length. For heavy materials, like lead, the value of λ_I decreases with increasing energy of the shower, reaching a value which corresponds closely to the geometrical cross section for the most energetic showers. For light materials, like carbon, the empirical evidence is not yet definite (§ 3.3, Part III), but it seems probable that λ_I will prove to be somewhat larger than the value corresponding to the geometrical cross section. Such a difference might be expected since it has been demonstrated in experiments on the absorption and scattering of 100 MEV. neutrons at Berkeley that the nuclei of light elements are more transparent than those of heavy elements.

The value of λ_A for air for penetrating showers is about 120 gm.cm^{-2}, and the value of λ_I is about 80 gm. cm^{-2} (allowing for transparency); thus one fast nucleon leads to approximately 1·5 nuclear collisions in air.

Besides the decrease in λ_I with increase in energy of the recorded shower there is also a noticeable decrease in λ_A with increase in energy. Thus for air λ_A for penetrating showers is about 120 gm. cm^{-2}, while for all stars it is about 140 gm. cm^{-2}. This difference is quite understandable on the picture of nucleon cascades (§ 4).

§ 2. THE PRIMARIES OF PENETRATING SHOWERS: THE A COMPONENT

By the A component is meant the nucleonic component which has sufficient energy to produce meson showers. This component is absorbed exponentially through much of the atmosphere with an attenuation length of 120 gm. cm^{-2}.

The A component consists of protons and neutrons in approximately equal numbers. At 11,000 ft. the ratio of neutral to primary radiation is $0{\cdot}83 \pm 0{\cdot}1$, which suggests that equilibrium has been established. The slight difference in the intensities may indicate a leakage of primary protons down through the atmosphere.

The interaction mean free paths of the charged and neutral primaries are the same.

2.1. *Exponential Absorption*

Heitler and Jánossy (1949) have given a general explanation of the exponential law of absorption on three basic assumptions : (i) One meson results from the collision of one nucelon with another; (ii) the probability $w(\epsilon, E)\,d\epsilon$ for an energy loss ϵ by a fast nucleon of energy E colliding with a nucleon at rest is given by the formula

$$w(\epsilon, E)\,d\epsilon = w(\epsilon/E)\,d\epsilon/E. \qquad \ldots\ldots(2.11)$$

The energy ϵ is the energy of the meson plus that of the recoil nucleon; (iii) the primary spectrum follows a power law. Each nucleon is assumed capable of producing two showers, and the following fluctuations are taken into account: (*a*) the number of nuclei 'hit', i.e. which have one primary particle passing through the geometrical cross section, follows a Poisson distribution; (*b*) the path length through the nucleus is determined by its spherical shape; (*c*) several nucleons may be hit when a primary particle passes through a nucleus, but the probability of an energy loss ϵ is given by formula (2.11) for each collision.

Assuming a reasonable index for the spectrum power law, the problem is soluble without further specification of $w(\epsilon/E)\,d\epsilon/E$. The result is that the spectrum of the primaries is reproduced for any thickness of absorber and the total intensity decreases exponentially with a mean free path which depends on one constant only. This constant can be determined and gives a reasonable figure for the mean free path.

2.2. *Meson Production*

The meson field of a nucleon can be compared in some respects to the electromagnetic field, the mesons of the meson field being analogous to the photons of the electromagnetic field. The cross section for the emission of photons in the electromagnetic case can be calculated using Born's approximation. This cross section diverges at low energies, leading to the emission of a large number of photons

of very low frequencies, but at high energies it converges and, on the average, only one photon is emitted per collision. In complete analogy the scattering of a fast nucleon by the meson field of another nucleon should lead to the emission of a meson. However, the meson field differs from the electromagnetic field in that mesons have electric charge and spin. The cross section in this case diverges at high energies.

Heisenberg (1949) has suggested that just as the multiple emission of photons occurs at low energies so several mesons may be produced in each nucleon–nucleon collision, the total number being limited by the conservation laws. Heitler has suggested that the spin and charge of the meson, which according to Heisenberg lead to multiple production, give rise to a strong damping term in the cross section and so prevent multiple production at high eneriges. According to Heitler only one meson is usually produced per nucleon–nucleon collision, the probability of multiple processes being about ten times less.

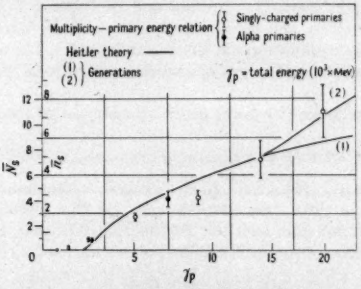


Figure 27. Theoretical and experimental results for the total number of mesons $\bar{N}_s$ and the number of charged mesons $\bar{n}_s$ as a function of the total energy γ_p of the primary particle. The theoretical curves are taken from the Heitler plural theory of meson production.
(*After* Davies, Franzinetti and Perkins, private communication)

Davies, Franzinetti and Perkins (private communication) have recently determined the energies of the primaries and secondaries of meson showers and they have examined their results in the light of the Heitler and Heisenberg theories. In Figure 27 the multiplicity of meson production is plotted against the total energy of the primary γ_p, the symbol $\bar{n}_s$ referring to shower mesons and $\bar{N}_s$ to all mesons. The full lines give the variation to be expected on Heitler's theory, curve (1) showing the effect of single nucleon–nucleon collisions of the same primary and curve (2) the same curve with the additional production of mesons from collisions of the first generation of knock-on nucleons. It can be seen that the theoretical and experimental results agree well. Heitler and Jánossy (1949, 1950) have extended their work to very large showers and can even account for large showers of the type observed by Kaplon, Bradt and Peters (Plate VI), in which as many as 56 shower particles were produced by an α-particle.

In Figure 28 the multiplicity–energy relation is plotted and compared with Heisenberg's theory, the various curves corresponding to different values of the constant K, the degree of inelasticity of the collision. This coefficient comes into

the expression for the energy ϵ available for all mesons in the centre-of-mass system as follows: $\epsilon = 2K(\gamma_u - 1)M_p/M_\pi$, where $\gamma_u = [(\gamma_p + 1)/2]^{1/2}$ and M_p and M_π are the proton and π-meson masses. The calculations of Heisenberg have been extended by Davies, Franzinetti and Perkins to include secondary generation, and with this addition the theory accounts well for the experimental results. These authors can also account for the observed number of mesons emitted in the backward direction in the laboratory system provided secondary production takes place.

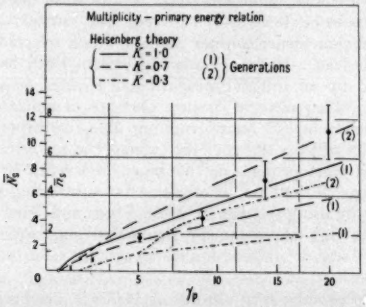


Figure 28. Theoretical and experimental results for the total number of mesons $\overline{N_s}$ and the number of charged mesons $\overline{n_s}$ as a function of the total energy γ_p of the primary particle. The theoretical curves are taken from the Heisenberg multiple theory of meson production. (*After* Davies, Franzinetti and Perkins, private communication)

Thus satisfactory agreement can be obtained equally well on both theories. This dual interpretation of the events follows from the fact that in both theories the main properties of penetrating showers depend on the conservation laws. The shower size is limited above all by the total energy available, and the angular distribution is largely determined by the conservation of momentum. Evidence for or against multiple processes must therefore await experiments in light nuclei, preferably hydrogen.

§ 3. SINGLE PROTONS

The protons belonging to type A which have been considered so far are of very high energy, a fact which is proved by the absence of a latitude effect for penetrating showers. As already pointed out, even the simplest penetrating shower set requires incident particles of energies greater than 10^{10} ev. Below this energy there exists in the atmosphere a continuous spectrum of protons the more energetic of which produce small penetrating showers. We proceed now to discuss the evidence for these protons. The state of our knowledge in this subject up to 1948 has been well summarized by Rossi.

At sea level most of these protons have energies of less than 5×10^9 ev.: they are therefore secondary to the primary cosmic radiation. They are probably mainly Heisenberg knock-on protons from stars formed in the atmosphere.

The spectrum of protons in the energy range 50–1,000 MEV., formed in large stars in the photographic emulsion at very great altitudes, has been discussed in a previous section (§ 2.2, Part II).

The two main methods which have been employed for the detection and examination of single protons in the atmosphere are the cloud chamber and the magnetic spectrograph. Below a momentum of 700 Mev/c. the cloud chamber can be used to identify protons directly, but above this momentum identification is indirect, involving an examination of the form of the spectrum and a knowledge of the sign of the particles. Rochester and Bound (1940) have made observations at sea level with a cloud chamber on protons in the momentum range $(3{\cdot}9-5{\cdot}5)\times10^8$ ev/c., and from their results Rossi (1948) has estimated the flux of particles to be $(0{\cdot}5\pm0{\cdot}17)\times10^{-6}$ cm^{-2} sec^{-1} sterad^{-1}. Rochester and Bound used an anticoincidence counter method which selected protons of less than 2 cm. range in lead. Hodson, Loria and Ryder (1950) have used a cloud chamber triggered by an internal proportional counter sensitive to heavily ionizing particles. They observe 26 slow particles of ionization greater than three times minimum in 727 hours' running time, corresponding to a total intensity of particles of $1{\cdot}5\times10^{-6}$ cm^{-2} sec^{-1} sterad^{-1} at sea level. The observed particles consist of slow protons in the range $(4{\cdot}3–6{\cdot}0)\times10^8$ ev/c. and slow mesons in the range $(9{\cdot}2–10{\cdot}1)\times10^7$ ev/c. After subtracting the contribution from slow mesons by using the data of Rossi, Sands and Sard (1947), Hodson *et al.* estimate the flux of proton in the stated momentum range to be $0{\cdot}7\times10^{-6}$ cm^{-2} sec^{-1} sterad^{-1}, in good agreement with the value found by Rochester and Bound.

The intensity of protons at an altitude of 34,000 ft. has been determined by Adams, Anderson, Lloyd, Rau and Saxena (1948). These workers observe single particles in a cloud chamber operated in a magnetic field. The momentum spectrum of the positive particles is found to be very different from the negative particles, and the difference is attributed to single protons. The observations cover the momentum range $(4–30)\times10^8$ ev/c. The flux of proton of momentum less than 10^9 ev/c. is about 10^{-2} cm^{-2} sec^{-1} sterad^{-1} and of momentum greater than 10^9 ev/c. is $8{\cdot}5\times10^{-3}$ cm^{-2} sec^{-1} sterad^{-1}. The proton intensity at 30,000 ft. is about 20% of the intensity of the meson component.

The magnetic spectrograph has been applied to this problem at sea level by Mylroi and Wilson (1951) and at an altitude of 11,000 ft. by Alikhanyan *et al.* (1948). In these experiments the absorption in lead of the particles of known sign and momenta which emerge from a magnetic spectrograph is determined. It is found that almost all of the particles which are anomalously absorbed are positive, and these particles are interpreted as protons.

The differential momentum spectrum observed by Mylroi and Wilson is shown in Figure 29. It can be represented by the formula $N(p)\,dp = Kp^{-2{\cdot}8}\,dp$ where $K = 0{\cdot}9\times10^{-4}$ cm^{-2} sec^{-1} sterad^{-1} if p and dp are measured in units of 10^9 ev/c. The low momentum cut-off of the spectrograph is 7×10^8 ev/c. The total flux above 10^9 ev/c. is $0{\cdot}50\times10^{-4}$ cm^{-2} sec^{-1} sterad^{-1}. Most of the particles are collimated in a vertical direction. The point obtained by Rochester and Bound is also shown on the diagram. The position of this point is quite consistent with the others when it is recalled that a proton of momentum 450 Mev/c. has a range of only 130 gm. cm^{-2} of air.

On comparing the results of Mylroi and Wilson at sea level and Adams *et al.* at 30,000 ft. the remarkable fact emerges that the form of the spectrum is the same but the intensity is very much reduced at sea level. Above 10^9 ev/c. the

ratio of the intensities is 1 : 170. This attenuation corresponds to a length of 140 ± 10 gm. cm^{-2}. A value consistent with this length is also found for the results of Alikhanyan *et al.* At an altitude of 11,000 ft. the intensity is 15·4 times the sea-level intensity and the flux of protons above 10^9 ev/c. is $7{\cdot}7 \times 10^{-4}$ cm^{-2} sec^{-1} sterad^{-1}.

The constant form of the spectrum over a thickness of two-thirds of the atmosphere can be interpreted on the picture of nucleon cascades as follows. Below 30,000 ft. there exists in the atmosphere a continuous spectrum of N-rays

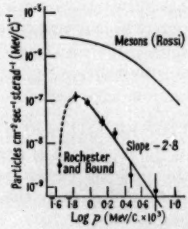


Figure 29. The differential momentum spectrum of single protons at sea level. (*After* Mylroi and Wilson, private communication)

which consist of protons and neutrons. As the protons penetrate the atmosphere, further protons are produced to replace those absorbed. There is a general reduction of $1 : e$ in a length of 140 gm. cm^{-2}, but a constant fraction of the particles which are absorbed is replaced by higher energy primaries for which the attenuation length is 120 gm. cm^{-2}. The larger attenuation length (i.e. 140 instead of 120 gm. cm^{-2}) is to be expected on this picture.

§ 4. THE PRIMARIES OF EVAPORATION STARS: THE B COMPONENT

B-type nucleons are those which produce evaporation stars, and it has already been pointed out that these are mainly neutrons. B-type protons are likely to be very rare. These nucleons are almost certainly secondary to and in equilibrium with a component which we shall term the A component, although it is intended here that this term shall include all fast nucleons which give rise to B-type nucleons. Following Fujimoto and Hayakawa (1949), the diffusion equations for the two components can be written

$$-\frac{dA(x)}{dx} = \frac{1}{\lambda_A} A(x) \quad \ldots\ldots(4.1), \qquad -\frac{dB(x)}{dx} = \frac{1}{\lambda_A} B(x) - \frac{k\nu}{\lambda_A} A(x), \quad \ldots\ldots(4.2)$$

where $A(x)$ and $B(x)$ represent the intensities of the A and B components at a depth x gm. cm^{-2} below the top of the atmosphere. For simplicity the same attenuation length is used for each component. The factor k gives the number of nuclear collisions in the length λ_A and is approximately 2. The quantity ν represents the average number of B-type nucleons produced at a collision. Solving these equations with the initial conditions $A(0) = a$ and $B(0) = 0$, and

noting that the B-type stars are formed by nucleons coming from all directions, Fujimoto and Hayakawa find

$$A(x) = a\exp(-x/\lambda_A), \quad \ldots\ldots(4.3), \qquad B(x) = 2\pi k\nu a(x/\lambda_A)[E_i(-x/\lambda_A)], \quad \ldots\ldots(4.4)$$

where the expression $E_i(-x/\lambda_A)$ is the Gross integral. Equation (4.4) reduces to an exponential, i.e. $b\exp(x/132)$ for $x \geqslant 200$ gm. cm^{-2} and $\lambda_A = 125$ gm. cm^{-2}. It fits the experimental results well. This equation shows a maximum at about 50 gm. cm^{-2}. The existence of such a maximum has already been noted (§ 5.2, Part III) but its exact position is not yet known.

G. P. Thomson (1949) has shown theoretically that an exponential law of absorption of the B component is to be expected if it is assumed that a fast neutron loses energy in a number of steps. It is also necessary that the energy spectrum should retain a constant form with depth.

§ 5. THE CORRELATION OF DIFFERENT TYPES OF STARS

We proceed now to discuss the correlation of penetrating showers and B-type stars. This can best be done by using the photographic emulsion data, since these data give the most complete information about the nature and absolute rates of formation of the different types of stars.

We will first discuss the correlation between the flux of primaries and the rate of showers with different numbers of relativistic particles. According to Camerini *et al.* (1949), the attenuation length of the primaries of stars with relativistic secondaries in air is 135 gm. cm^{-2}. Assuming that the primary flux at the top of the atmosphere is approximately 0·10 particle cm^{-2} sec^{-1} sterad^{-1}, the flux of particles at 11,000 ft. with energies above the geomagnetic cut-off at the Jungfraujoch (i.e. 3,500 Mev.) can be calculated. With the Gross transformation the result is 50 particles cm^{-2} day^{-1}. Primaries of energy greater than 3,500 Mev. can produce stars with at least one relativistic secondary particle (see Figure 21).

Now the flux of primaries at the Jungfraujoch can also be calculated from the observations of the number of stars of different types observed in the photographic emulsion. Using the data of Brown *et al.*, and assuming a cross section for star production equal to the geometrical cross section, the flux of primaries producing stars with $n_s \geqslant 2$ is 14 particles cm^{-2} day^{-1}, and with $n_s \geqslant 1$ is 40 particles cm^{-2} day^{-1}. The agreement between the two estimates of the flux of primaries for $n_s \geqslant 1$ is quite reasonable in view of the assumptions which have been made, and leads to the conclusion that the primaries of these showers are the secondary products of the primary flux of particles at the top of the atmosphere.

We will next discuss the relative numbers of stars of different types in the emulsion. Brown *et al.* have found at the Jungfraujoch that the rate of stars of the types (0n) : (0p + 1p + 2p + ... 1n + 2n + 3n + ...) is 3·2 : 1. Thus in a photographic plate at the Jungfraujoch approximately three evaporation stars (i.e. of type 0n) are in equilibrium with one event with at least one relativistic particle. It is reasonable to assume that the evaporation stars are secondary to stars with relativistic particles. The evaporation stars in a photographic plate have, however, come from events produced one attenuation length of air above the plates. One relativistic event at this (upper) level in the atmosphere will be in equilibrium with a certain number n evaporation stars at the plate. The rate of

relativistic events at the plate will, however, be only $1/e$ times the rate at the upper level. Thus the ratio of evaporation stars to relativistic events at the plate will will be ne : 1. Since this ratio is found experimentally to be 3:1, n is approximately equal to unity.*

§6. SLOW PROTONS AND SLOW NEUTRONS: THE C COMPONENT

6.1. *Slow Protons*

By the C component is meant the ultimate product of stars—that is, mainly slow protons and slow neutrons—of such low energies that they cannot produce stars. The average energy of an evaporation proton is about 15 MeV.

The investigation of slow protons has been carried out with the cloud chamber and the photographic plate. Excellent surveys up to 1948 have already been given by Montgomery (1949) and Rossi (1948).

The variation with altitude of these slow protons has been measured with the photographic plate technique by Widhalm (1940) and Perkins (1947b). These authors found that the rate varied with altitude in the same way as for evaporation stars, a result which shows that these particles are secondary to stars. Perkins also found an attenuation length for single tracks in lead of 320 gm/cm², a value which is again consistent with their connection with stars.

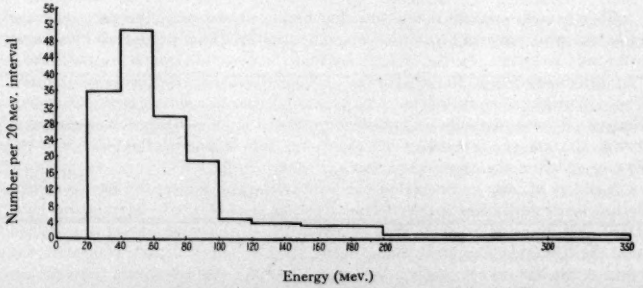


Figure 30. The energy distribution of all protons, whether produced in stars or singly, in the electron-sensitive photographic emulsion, at an altitude of 3,650 m. (*After* Lattimore 1950)

Perkins and Widhalm both thought that these single tracks came from stars formed in matter surrounding the plates, but Lattimore (1949, 1950) in two recent papers has shown convincingly that they arise mainly in stars in the emulsion. In his first paper Lattimore investigated with Ilford C2 plates the angular distribution, the frequency and the absorption of single proton tracks. The angular distribution suggests that the protons consist of two groups, one having an isotropic distribution and the other collimated downwards. At an altitude of 3,650 m. only about 10% of the observed tracks lie in the first group, and they can be accounted for by stars formed in the emulsion and glass of the plates. The total flux of single protons is 16·5 particles/cm²/day. There is little change

* As this Report was going to press we received a communication from Dr. H. Messel in which the problem of nucleon cascades is treated theoretically. The predictions of the theory seem to agree well with the experimental results.

in the number of single tracks with the thickness of glass over and around the plates or with small thicknesses of ice. This suggests that the tracks cannot be due to proton knock-ons from neutrons. Bagge has suggested that the tracks are due to one-pronged stars with an energy distribution similar to the many-pronged stars; however, observations by Lattimore in C2 plates seemed to indicate that single-pronged stars could only account for some 30% of the tracks.

In his later work Lattimore has used Ilford G5 200 μ plates (exposed at 3,650 m.). The emulsion has been searched twice with a 95 × oil-immersion objective, and 74 single protons of energy greater than 50 MeV. have been found which start in the emulsion. This corresponds to a rate of 25/cm³/day. In addition, many protons have been observed to be emitted from stars which could not be detected in the C2 plates. The rate of stars with two or more prongs is 23/cm³/day, giving a total rate of production of protons of energy greater than 30 MeV. of 28/cm³/day. Thus the total rate of production of fast protons from both causes is about 53/cm³/day. The energy distribution of these protons is shown in Figure 30. With these data it is now possible to account for all the single tracks. It is concluded that the agency which produces stars also produces single protons.

6.2. *Slow Neutrons*

(*a*) *Slow neutrons in the atmosphere.*

Slow neutrons can be observed either by placing a piece of paraffin wax over a photographic plate and counting the number of knock-on protons or by a boron trifluoride counter. In the second method the slow neutron is captured by a ^{10}B nucleus, leading to the production of ^{7}Li and an α-particle of energy 1·89 MeV. The α-particle is recorded by a proportional counter or ionization chamber. Neither of these methods is very efficient, and a large correction is necessary to obtain the correct intensity. Furthermore, the second method involves the slowing down of the neutrons to thermal energies.

A survey of early work on slow neutrons in the atmosphere has been given by Bethe, Korff and Placzek (1940) and by Flügge (1943). It has already been stated that evaporation stars show a maximum in the atmosphere, as expected from their secondary origin. Since nearly all slow neutrons come from stars, the neutron rate should also show a maximum, but at a slightly greater depth in the atmosphere. The distribution for slow neutrons has been calculated theoretically by Bethe *et al.* and Flügge who find that the maximum should occur at a depth of 130 gm/cm².

Recent experiments on slow neutrons in the atmosphere have been carried out by Simpson (1948) and Ladenburg and Yuan (1949). The former used a proportional boron trifluoride counter covered with paraffin wax and the latter a boron trifluoride counter surrounded by a suitable moderator. Simpson's experiment consisted in sending up a counter surrounded by paraffin wax covered on the outside by a thin layer of cadmium. Cadmium absorbs all thermal neutrons, and Simpson's apparatus records only neutrons above an energy of 0·4 ev. Ladenburg and Yuan compare the counting rate of two identical counters filled with boron trifluoride, one being surrounded by tin foil and the other by cadmium foil. The counter surrounded with cadmium gives the background rate, and the difference in the rates of the two counters is a measure of the flux of thermal neutrons. It is found that at a latitude of 52° N. up to an altitude of 50,000 ft., the rate of thermal neutrons increases corresponding to an

attenuation length of 160 gm. cm^{-2}. Above the level corresponding to a pressure of 8 cm. Hg, however, the counting rate decreases rapidly, reaching a value of one-quarter of the maximum rate at a pressure of 0·75 cm. Hg. The position of the maximum agrees with the predictions of Flügge.

At 53° N. Simpson also finds an attenuation length of 160 gm/cm^2. The same value is to be expected, as in each method a constant fraction of the slow neutrons is recorded. The value of 160 gm. cm^{-2} is slightly longer than the attenuation length for evaporation stars (i.e. 140 gm. cm^{-2}), but the difference is consistent with the secondary nature of neutrons.

The latitude effect for slow neutrons has been determined by Yuan (1949) and Simpson (1949). Yuan finds that the ratio of neutron intensities between 50° N. and 20° N. at an altitude of 30,000 ft. is 3·5. Simpson finds an effect between 53° N. and 0° N. of 3·5 at a depth in the atmosphere of 200 gm. cm^{-2} and 2·3 at a depth of 600 gm. cm^{-2}. These results show that the latitude effects for slow neutrons are the same as for evaporation stars, a fact which is again in line with the generic connection of neutrons and stars.

The value of λ_A for slow neutrons has been determined at various latitudes by Simpson (1949) and his results are as follows: at 53° N., 40° N. and at 0° and 19° N. the values are 160 gm. cm^{-2}, 185 gm. cm^{-2} and 200 gm. cm^{-2} respectively. The increase in the value of λ_A with decrease of latitude is very marked and is a consequence of the increase in mean energy of cosmic-ray primaries. Thus the primary cosmic-ray particles at the equator produce a longer train of nuclear interactions than the primaries at 53° N., and at each interaction slow neutrons are emitted.

It is difficult to estimate the total flux of slow neutrons at different altitudes because of the difficulty of assessing the efficiency of the recording instruments. According to Bethe *et al.* (1940) the rate of production of neutrons in the atmosphere is 0·05 neutrons/gm. at a residual pressure of about 100 gm. cm^{-2}. Assuming an attenuation length of 150 gm. cm^{-2} Flügge concludes that the number of slow neutrons produced in a column of air of one square centimetre cross section throughout the atmosphere is 15, which corresponds to about 50 neutrons per primary cosmic-ray particle. This figure may be in error by a factor of two. Even so, it is consistent with the general scheme we have outlined, using the fact that about 1·5 evaporation stars are in equilibrium with each shower containing one or more than one relativistic particle, and also that 1·5 showers with one relativistic particle are in equilibrium with each penetrating shower. Thus each penetrating shower will produce a series of about five nuclear interactions, in each of which about four slow neutrons are emitted. To obtain a value of 50 slow neutrons per primary would then require the production of about two penetrating showers per primary. This is a reasonable figure.

An unusual increase in the intensity of slow neutrons was observed by Adams (1950) at Manchester, who found a fivefold increase in rate over a short period of a few hours on 19th November 1949. The increase in the penetrating component over the same period was only 12%. No change in cosmic-ray intensity was observed at the equator. The effect has been interpreted by Adams and Braddick (1950) as due to an increase in the intensity of primary cosmic rays in the momentum range (3–19) $\times 10^9$ ev/c. which gives rise to an appreciable number of slow neutrons at sea level but leads to the production of very few μ-mesons observable at sea level. In line with this explanation is the observation that the

latitude effect of μ-mesons between the equator and 50° N. is very small compared with the effect for slow neutrons.

(b) Slow neutrons produced locally in condensed matter.

Tongiorgi (1949) and Montgomery and Tobey (1949) have carried out experiments on the production of slow neutrons in absorbers. In these experiments the recording unit, which consists of systems of boron trifluoride counters surrounded by paraffin, is placed below or within various absorbers. Clearly the counters will detect slow neutrons from two sources: the absorber, and the matter outside the apparatus. The intensity of neutrons from outside the apparatus is determined by replacing the absorber by borax; this quantity is subtracted from the recorded rate. It is then found that the effect of a dense absorber like lead is to produce a marked increase in the rate of neutrons at the boron trifluoride counters. Moreover, the rate of increase of events recorded by systems of boron trifluoride counters with altitude gives an attenuation length which is closely similar to the value for stars. This result may be taken to indicate that the slow neutrons produced in absorbers come from the same primaries as those which produce stars.

By observing the coincidence of boron trifluoride counters Tongiorgi finds that the average multiplicity per coincidence, that is per star, is 8 in lead and 1·6 in paraffin. These values are somewhat uncertain because of uncertainties in the calibration of the apparatus and the difficulty of eliminating the effects of a number of contemporary stars. Bearing this in mind, the result is consistent with the multiplicity observed in photographic plates where in silver and bromine the number of neutrons emitted per star is about 5.

Tongiorgi has also shown that the relative rates of production of slow neutrons in various substances correspond to a cross section for production proportional to $A^{2/3}$. Allowance is made in this calculation for the variation in average multiplicity of the stars formed in different materials. This result is also in agreement with the hypothesis that the slow neutrons are the secondary products of stars.

The association of slow neutrons with penetrating showers was first shown by Cocconi, Tongiorgi and Greisen (1948); a more detailed study of this problem has been made by Cocconi, Tongiorgi and Widgoff (1950). In this investigation observations are made of the coincidences of an array of boron trifluoride counters with a Geiger counter set sensitive to showers with at least one penetrating secondary particle (see Figure 31). The discharges of the boron trifluoride counters are recorded individually so that not only the number but also the grouping of the counters is known. It is found that showers produced by both charged and uncharged primary particles emit large numbers of slow neutrons isotropically. The number of neutrons produced increases rapidly with the increase in the thickness of the absorber Σ. Some typical results are given in Table 13. It is considered that the relative values of the mean multiplicities are correct to 20%, but that the actual numbers may be in error by a factor of two.

It is seen that the number of neutrons increases rapidly with the thickness of the absorber. This is interpreted by Cocconi *et al.* as evidence for an increase in the number of nuclear interactions with the thickness of the absorber.

The size–frequency distributions for the different thicknesses are given in column (4) of the table. It may be noted that each is of exponential form similar to the form of the size–frequency distributions for charged particles from

stars (see equation (1.21), Part II). The photographic plate data will probably correspond most nearly to the observations on the smallest thicknesses of absorber where successive production of neutrons is least. Even in this case, however, it is seen that the exponent for charged particles is at least three times the value

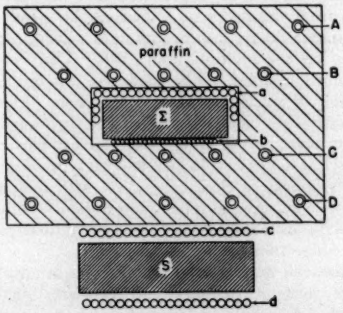


Figure 31. Arrangement used for the detection of neutrons in penetrating showers. Trays a, b, c and d consist of G–M counters and trays A, B, C and D of BF_3 proportional counters. The events studied are showers produced by ionizing or non-ionizing particles in the absorber Σ. Production by an ionizing particle is shown by the discharge of the counter in tray a.

The master pulse consists of a coincidence bc together with the discharge of one BF_3 counter in the time interval 5–15 μsec. after the coincidence bc. All the counters are hodoscoped, and thus the type of event can be ascertained. The number and grouping of the BF_3 counters which are set off are also observed; from these data the number of neutrons produced can be estimated. (*After* Cocconi, Tongiorgi and Widgoff 1950)

for the neutrons. The difference is ascribed by Cocconi *et al.* to the absence of a Coulomb barrier for neutrons and the selection of high-energy events by their counter set. A better comparison might be made by comparing the size–frequency distribution of events of the type 1n, 1p, 2n, 2p, etc., with the neutron spectra.

Table 13. Average Multiplicity and the Size–Frequency Distributions of Slow Neutrons emitted in Penetrating Showers at 3,260 m.

(1)	(2)	(3)	(4)
Lead	4·50	50	$10 \exp(-0{\cdot}02v)$
,,	1·00	17	$30 \exp(-0{\cdot}06v)$
,,	0·25	10	$30 \exp(-0{\cdot}1v)$
Aluminium	4·75	6	$50 \exp(-0{\cdot}17v)$

(1) Absorbers Σ. (2) Thickness of absorber Σ (in.). (3) Average neutron multiplicity. (4) Size–frequency distribution (frequency of events with v neutrons per hr.).

V. CONCLUSION

The general picture which emerges from this review of the nuclear interactions of cosmic rays is one of rapid progress in the accumulation of empirical facts but little advance in our fundamental theoretical concepts. The use of powerful new techniques, and the further development of well-established techniques, have led in the short space of the past few years to the discovery of the π-mesons and to the indication of the presence of other types of unstable particle. The main

process of meson formation is certainly π-meson production in stars produced near the top of the atmosphere. Results are just beginning to appear on the multiplicity of formation of these mesons as a function of star size and energy of the primary particle. Moreover, it is clear that μ-mesons are secondary to these primary π-mesons, and it is probable that the soft component in the atmosphere arises from the decay of π^0-mesons also produced in the primary process. The π^0-meson decays into two photons which produce large electron–photon showers by cascade multiplication. What is still not clear is the exact mechanism of meson formation. The answer to this question awaits an exact theoretical treatment of nucleon–nucleon interaction. Experiments on penetrating showers with light materials, especially hydrogen, may give much information on this point. It is still not clear what part the heavy, unstable particles play in nuclear forces and, in fact, in what type of interaction they arise. The empirical facts indicate a close connection with meson showers, but it is possible that they arise in a special type of interaction. Information is required on the types of heavy, unstable particle which exist, their properties and interrelation, and their connection with other mesons.

The general scheme of nuclear interactions in the atmosphere, leading to the formation of huge nucleon cascades and extensive showers, is probably qualitatively correct. A clearer understanding of the meaning of interaction and attenuation lengths and of their connection one with another is gradually emerging. A more precise knowledge of their values in air and condensed matter is still needed. The picture of nucleon cascades has still to be extended in detail to condensed matter, where the effects of π-mesons, and possibly μ-mesons, may be important. The difficulties which arise in this type of analysis come not only from the complication of the phenomena but also from the absence of precise knowledge of the behaviour of the different sorts of particles in different energy ranges. Below 10^9 ev. information is being rapidly acquired of the cross sections for nuclear processes for the different sorts of elementary particle, but above 10^9 ev. it is not yet possible to distinguish empirically one elementary charged particle from another. It is thus difficult to assess the contribution of the different kinds of elementary particle. One would like to know, in particular, how the cross sections for nuclear processes of the π-meson and the proton vary with energy. Experiments with cloud chambers and electron-sensitive emulsions in high magnetic fields might give some relevant information.

Near the top of the atmosphere the heavy nuclei probably play as important a role as protons in the formation of large stars. The interaction of these nuclei with matter, and their formation in stars will no doubt be studied intensively in the near future.

The empirical facts relating to evaporation stars and the statistical theory of the evaporation of excited nuclei are now well established. Details of the process of formation of these stars still need to be worked out, for example the large emission of low-energy α-particles and the emission of high-energy nuclear fragments. Experiments in stars in elements of high atomic weight are still desirable. A further question is how meson formation is connected with star size. Preliminary results seem to indicate that meson multiplicity is only slightly connected with the numbers of grey and black tracks. These facts need to be established, and they will have to be incorporated in any theory of meson formation.

Much remains then to be done, but the prospects are exciting. If the progress of the past few years can be maintained the coming years will see further rapid advances in this field of research, and a solution of many of the problems which still baffle the nuclear physicist.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the helpful criticism of Dr. J. G. Wilson and Dr. C. C. Butler, who kindly read the Report in manuscript form.

We are grateful to the many authors who have sent us information or copies of their papers in advance of publication. These papers and photographs and diagrams are acknowledged in the text.

The reviews of Professor Rossi (1948, 1949), and the Report of Professor Powell (1950) have been freely used in the preparation of this Report.

Finally, we wish to express out thanks for the permission granted by various authors and publishing bodies for the reproduction of diagrams and photographs.

REFERENCES

ADAMS, N., 1950, *Phil. Mag.*, **41**, 503.
ADAMS, N., and BRADDICK, H. J. J., 1950, *Phil. Mag.*, **41**, 505.
ADAMS, R. V., ANDERSON, C. D., LLOYD, P, E., RAU, R. R., and SAXENA, R. C., 1948, *Rev. Mod. Phys.*, **20**, 234.
ADDARIO, M. M., and TAMBURINO, S., 1949, *Nuovo Cim.*, **6** (suppl.), 429.
ALIKHANYAN, A. I., ALIKHANOV, A. I., and WEISSENBERG, A., 1948, *J. Exp. Theor. Phys. U.S.S.R.*, **18**, No. 8.
APPAPILLAI, V., and MAILVAGANAM, A. W., 1948, *Nature, Lond.*, **162**, 1792; 1950, *Proc. Phys. Soc.* A, **63**, 856.
ARMENTEROS, R., BARKER, K. H., BUTLER, C. C., CACHON, A., and CHAPMAN, A. H., 1951, *Nature, Lond.*, in the press.
BAGGE, E., 1941, *Ann. Phys., Lpz.*, **39**, 512, 533; 1943, *Cosmic Radiation* (New York: reprinted by Dover Publications), p. 128.
BALDWIN, G. C., and KLAIBER, G. S., 1946, *Phys. Rev.*, **70**, 259.
BARKER, K. H., and BUTLER, C. C., 1951, *Proc. Phys. Soc.* A, **64**, 4.
BARTON, J. C., GEORGE E. P., and JASON, A. C., 1951, *Proc. Phys. Soc.* A, **64**, 175.
BERNARDINI, G., CORTINI, G., and MANFREDINI, A., 1949, *Phys. Rev.*, **76**, 1792; 1950, *Ibid.*, **79**, 952.
BETHE, H. A., 1937, *Rev. Mod. Phys.*, **9**, 69; 1940, *Phys. Rev.*, **57**, 1125.
BETHE, H. A., KORFF, S. A., and PLACZEK, G., 1940, *Phys. Rev.*, **57**, 573.
BJORKLAND, R., CRANDALL, W. E., MOYER, B. J., and YORK, H. T., 1950, *Phys. Rev.*, **77**, 213.
BLAU, M., and WAMBACHER, H., 1937, *S.B. Akad. Wiss. Wien*, IIA, **146**, 623.
BOHR, N., and KALCKAR, F., 1937, *Mat.-Fys. Medd. K. Danske Vid. Selskab.*, **24**, 10.
BONETTI, A., and DILWORTH, C., 1949, *Phil. Mag.*, **40**, 947.
BRADT, H. L., and PETERS, B., 1949, *Cosmic Radiation*: *The Colston Papers* (London: Butterworth), p. 5; 1950, *Phys. Rev.*, **77**, 54.
BRIDGE, H., and HAZEN, W. E., 1948, *Phys. Rev.*, **74**, 579.
BRIDGE, H., HAZEN, W. E., ROSSI, B., and WILLIAMS, R. W., 1948, *Phys. Rev.*, **74**, 1083.
BRIDGE, H., and ROSSI, B., 1949, *Phys. Rev.*, **75**, 810.
BRIDGE, H., ROSSI, B., and WILLIAMS, R. W., 1947, *Phys. Rev.*, **72**, 257.
BROADBENT, D., and JÁNOSSY, L., 1947, *Proc. Roy. Soc.* A, **191**, 817.
BROWN, R. H., CAMERINI, U., FOWLER, P. H., HEITLER, H., KING, D. T., and POWELL, C. F., 1949, *Phil. Mag.*, **40**, 862.
BROWN, W. W., and MCKAY, A. S., 1949, *Phys. Rev.*, **76**, 1034; 1950, *Ibid.*, **77**, 342.
BUTLER, C. C., ROSSER, W. G. V., and BARKER, K. H., 1950, *Proc. Phys. Soc.* A, **63**, 145.
CAMERINI, U., COOR, T., DAVIES, J. H., FOWLER, P. H., LOCK, W. O., MUIRHEAD, H., and TOBIN, N., 1949, *Phil. Mag.*, **40**, 1073.

CAMERINI, U., FOWLER, P. H., LOCK, W. O., and MUIRHEAD, H., 1950, *Phil. Mag.*, **41**, 413.
CARLSON, A. G., HOOPER, J. E., and KING, D. T., 1950, *Phil. Mag.*, **41**, 701.
CHAO, C. Y., 1949, *Phys. Rev.*, **75**, 581.
CHOWDHURI, B., 1950, *Proc. Phys. Soc.* A, **63**, 165.
CLEMENTEL, E., and PUPPI, G., 1949, *Nuovo Cim.*, **6**, 1.
COCCONI, G., 1949 a, *Phys. Rev.*, **75**, 1074; 1949 b, *Ibid.*, **76**, 984.
COCCONI, G., LOVERDO, A., and TONGIORGI, V., 1946, *Phys. Rev.*, **70**, 852.
COCCONI, G., TONGIORGI, V. C., and GREISEN, K., 1948, *Phys. Rev.*, **74**, 1867.
COCCONI, G., TONGIORGI, V. C., and WIDGOFF, M., 1950, *Phys. Rev.*, **79**, 768.
COOR, T., and SNOW, G. A., 1948, *Phys. Rev.*, **73**, 1252.
CORTINI, G., and MANFREDINI, A., 1949, *Nature, Lond.*, **163**, 992.
DALLAPORTA, N., MERLIN, M., and PUPPI, G., 1950, *Nuovo Cim.*, **7**, 99.
DAUDIN, J., 1942, *Thesis*, Paris; 1943, *Ann. Phys., Paris*, **18**, 217.
FERNBACH, S., SERBER, R., and TAYLOR, T. T., 1949, *Phys. Rev.*, **75**, 1352.
FLÜGGE, E., 1943, *Cosmic Radiation* (New York : reprinted by Dover Publications), Chap. 14.
FOWLER, P. H., 1950, *Phil. Mag.*, **40**, 169.
FREIER, P., NEY, E. P., and OPPENHEIMER, F., 1949, *Phys. Rev.*, **75**, 1451.
FRETTER, W., 1948, *Phys. Rev.*, **73**, 41; 1949 a, *Ibid.*, **76**, 511; 1949 b, *Amer. J. Phys.*, **17**, 148.
FUJIMOTO, Y., and HAYAKAWA, S., 1949, *Prog. Theor. Phys.*, **4**, 502.
FUJIMOTO, Y., HAYAKAWA, S., and YAMAGUCHI, Y., 1949, *Prog. Theor. Phys.*, **4**, 576.
FUJIMOTO, Y., and YAMAGUCHI, Y., 1949, *Phys. Rev.*, **75**, 1775.
FUSSEL, L., 1937 (referred to by EULER, H., and HEISENBERG, W., 1938, *Ergebn. Exakt. Naturw.*, **17**, 1.)
GARDNER, E., 1949, *Bull. Amer. Phys. Soc.*, Feb. 3.
GARDNER, E., and PETERSON, V., 1949, *Phys. Rev.*, **75**, 364.
GEORGE, E. P., 1949, *Nuovo Cim.*, **6** (suppl.), 531.
GEORGE, E. P., and EVANS, J., 1950, *Proc. Phys. Soc.* A, **63**, 1248; 1951, *Proc. Phys. Soc.* A, **64**, 193.
GEORGE, E. P., and JASON, A. C., 1948, *Nature, Lond.*, **161**, 218; 1949, *Proc. Phys. Soc.* A, **62**, 243; 1950, *Ibid.*, **63**, 1081.
GOLDSCHMIDT-CLERMONT, Y., 1950, *Nuovo Cim.*, **7**, 1.
GREISEN, K. I., 1942, *Phys. Rev.*, **61**, 212.
HARDING, J. B., 1949 a, *Phil. Mag.*, **40**, 530; 1949 b, *Nature, Lond.*, **163**, 440.
HARDING, J. B., LATTIMORE, S. A., and PERKINS, D. H., 1949, *Proc. Roy. Soc.* A, **196**, 325.
HARDING, J. B., LATTIMORE, S., LI, T. T., and PERKINS, D. H., 1949, *Nature, Lond.*, **163**, 320.
HAYAKAWA, S., 1949, *Phys. Rev.*, **75**, 1759.
HAZEN, W. E., 1944, *Phys. Rev.*, **65**, 67.
HEISENBERG, W., 1937, *Sachs Akad. Wiss., Lpz., Maths. Phys. Kl. Ber.*, **89**, 369; 1949, *Nature, Lond.*, **164**, 65.
HEITLER, W., and JÁNOSSY, J., 1949, *Proc. Phys. Soc.* A, **62**, 374; 1950, *Helv. Phys. Acta*, **23**, 417.
HODGSON, P. E., and PERKINS, D. H., 1949, *Nature, Lond.*, **163**, 439.
HODSON, A. L., LORIA, A., and RYDER, N. V., 1950, *Phil. Mag.*, **41**, 826.
HORNING, W., and BAUMHOFF, L., 1949, *Phys. Rev.*, **75**, 364.
HULSIZER, R. I., 1948, *Phys. Rev.*, **73**, 1252.
ISE, J., and FRETTER, W., 1949, *Phys. Rev.*, **76**, 933.
JÁNOSSY, L., 1942, *Proc. Roy. Soc.* A, **179**, 361; 1943, *Phys. Rev.*, **64**, 345; 1944, *Proc. Roy. Soc.* A, **183**, 190; 1948, *Cosmic Rays*, 1st Edition (Oxford: Clarendon Press), p. 358.
JÁNOSSY, J., and INGLEBY, P., 1940, *Nature, Lond.*, **145**, 511.
JÁNOSSY, L., and ROCHESTER, G. D., 1943, *Proc. Roy. Soc.* A, **182**, 180; 1944 a, *Ibid.*, **183**, 181; 1944 b, *Ibid.*, **183**, 186.
KAPLON, M. F., PETERS, B., and BRADT., H. L., 1949, *Phys. Rev.*, **76**, 1735.
LADENBURG, R., and YUAN, L. C. L., 1949, *Cosmic Radiation* : *The Colston Papers* (London : Butterworth), p. 35.
LATTIMORE, S., 1949, *Phil. Mag.*, **40**, 394; 1950, *Ibid.*, **41**, 819.

Le Couteur, K., 1950 a, *Proc. Phys. Soc.* A, **63,** 259; 1950 b, *Ibid.*, **63,** 498.
Leprince-Ringuet, L., Bousser, F., Hoang-Tchang-Fong, Jauneau, L., and Morellet, D., 1949, *Phys. Rev.*, **76,** 1273.
Lord, J. J., and Schein, M., 1950, *Phys. Rev.*, **77,** 19.
Lovati, A., Mura, A., Salvini, G., and Tagliaferri, G., 1950, *Nuovo Cim.*, **7,** 36.
Malaspina, L., Merlin, M., Pierucci, O., and Rostagni, A., 1950, *Nuovo Cim.*, **7,** 145.
McCusker, C. B. A., 1950, *Proc. Phys. Soc.* A, **63,** 1240.
Menon, M. G. K., Muirhead, H., and Rochat, O., 1950, *Phil. Mag.*, **41,** 583.
Mezzetti, L., and Querzoli, R., 1950, *Phys. Rev.*, **179,** 169.
Mitra, S. M., and Rosser, W. G. V., 1949, *Proc. Phys. Soc.* A, **62,** 364.
Montgomery, C. G., and Tobey, A. R., 1949, *Phys. Rev.*, **76,** 1478.
Montgomery, D. J. X., 1949, *Cosmic Ray Physics* (Princeton : University Press), p. 240.
Mylroi, M. G., and Wilson, J. G., 1951, *Proc. Phys. Soc.* A, **64,** 404.
Occhialini, G. P. S., and Cosyns, M. G. E., 1949, *Nuovo Cim.*, **6** (suppl.), 413
Page, N., 1950, *Proc. Phys. Soc.* A, **63,** 250.
Panofsky, W. K. H., Aamodt, L., and York, H. F., 1950, *Phys. Rev.*, **78,** 825.
Perkins, D. H., 1947 a, *Nature, Lond.*, **160,** 299; 1947 b, *Ibid.*, **160,** 707; 1949, *Phil. Mag.*, **40,** 601; 1950, *Ibid.*, **41,** 138.
Peterson, J. M., McMillan, E. M., and White, R. S., 1950, *Phys. Rev.*, **78,** 84.
Piccioni, O., 1950, *Phys. Rev.*, **77,** 1.
Powell, C. F., 1950, *Rep. Prog. Phys.*, **13** (London : Physical Society), p. 350.
Powell, W. M., 1946, *Phys. Rev.*, **69,** 385.
Rochester, G. D., 1946, *Proc. Roy. Soc.* A, **178,** 464.
Rochester, G. D., and Bound, M., 1940, *Nature, Lond.*, **146,** 745.
Rochester, G. D., and Butler, C. C., 1947, *Nature, Lond.*, **160,** 855; 1948, *Proc. Phys. Soc.* A, **61,** 535.
Rochester, G. D., Butler, C. C., and Runcorn, S. K., 1947, *Nature, Lond.*, **159,** 227.
Rosenbluth, M. N., 1949, *Phys. Rev.*, **75,** 532.
Rossi, B., 1948, *Rev. Mod. Phys.*, **20,** 537; 1949, *Proceedings of the Echo Lake Cosmic Ray Symposium* (Washington, D.C. : U.S. Office of Naval Research), p. 307.
Rossi, B., and Greisen, K., 1941, *Rev. Mod. Phys.*, **13,** 240.
Rossi, B., and Regener, V., 1940, *Phys. Rev.*, **58,** 837.
Rossi, B., Sands, M., and Sard, R. F., 1947, *Phys. Rev.*, **72,** 120.
Ruddlesden, S. N., and Clark, A. C., 1949, *Nature, Lond.*, **164,** 487.
Sala, U., and Wataghin, G., 1946, *Phys. Rev.*, **67,** 55.
Salant, S. O., 1949, *see under* Rossi, B., 1949, p. 173.
Sands, M., 1950, *Phys. Rev.*, **77,** 180.
Sard, R. D., Conforto, A. M., and Crouch, M. F., 1949, *Phys. Rev.*, **76,** 1134.
Seriff, A. J., Leighton, R. B., Hsaio, C., Cowan, E. W., and Anderson, C. D., 1950, *Phys. Rev.*, **78,** 290.
Simpson, J. A., 1948, *Phys. Rev.*, **74,** 1214; 1949, *see under* Rossi, B., 1949, p. 175.
Sitte, K., 1950 a, *Phys. Rev.*, **78,** 714; 1950 b, *Ibid.*, **78,** 721.
Sorensen, S. O. C., 1949, *Phil. Mag.*, **40,** 947.
Stetter, G., and Wambacher, H., 1939, *Phys. Z.*, **40,** 702.
Tamor, S., 1950, *Phys. Rev.*, **77,** 412.
Thomson, G. P., 1949, *Phil. Mag.*, **40,** 589.
Ticho, H. K., 1948, *Phys. Rev.*, **74,** 1337.
Tinlot, J., 1948 a, *Phys. Rev.*, **73,** 1476; 1948 b, *Ibid.*, **74,** 1197.
Tinlot, J., and Gregory, B., 1949, *Phys. Rev.*, **75,** 519.
Tongiorgi, V. C., 1949, *Phys. Rev.*, **76,** 517.
Vallarta, M. S., 1950, *Phys. Rev.*, **77,** 419.
Walker, W. D., 1950, *Phys. Rev.*, **77,** 686.
Walker, W. D., Walker, S. P., and Greisen, K., 1950, *Phys. Rev.*, **80,** 546.
Walsh T. G., and Piccioni, O., 1950, *Phys. Rev.*, **80,** 619.
Wataghin, G., de Santos, H., and Pompeia, P. A., 1940, *Phys. Rev.*, **57,** 61, 339.
Weisskopf, V., 1937, *Phys. Rev.*, **52,** 293.
Widhalm, A., 1940, *Z. Phys.*, **115,** 481.
Yagoda, H., 1949, *see under* Rossi, B., 1949, p. 169.
Yuan, L. C. L., 1949, *Phys. Rev.*, **76,** 1267.

NUCLEAR EFFECTS IN ATOMIC SPECTRA

By E. W. FOSTER

CONTENTS

	PAGE
§ 1. Introduction	288
§ 2. Isotope structure	289
§ 3. The nuclear mass effect	290
§ 4. The nuclear field effect	293
(i) Nature of the observational material	293
(ii) Assignment of isotope shifts to spectrum terms	295
(iii) Theory of isotope shift in heavy element spectra ('volume effect')	296
(iv) Comparisons of experimental data with volume effect calculations	299
(v) The effect of screening of the inner electrons by the optical electrons	304
(vi) Configuration perturbation	306
(vii) Second approximation in the calculation of the volume effect	306
(viii) The 'staggering' of the odd isotope centroids	307
§ 5. Nuclear moments	308
(i) The effect of finite distribution of nuclear charge	309
(ii) The effect of finite distribution of nuclear magnetic moment	311
(iii) The effect of the proton motion in hydrogen and deuterium	312
Conclusion	313
Acknowledgments	313
References	313

ABSTRACT. The report is concerned mainly with results of optical investigations of hyperfine structure, particularly of isotope structure. Recent work on the nuclear mass effect in He is described. Analysis of ZnI isotope shifts has shown that nuclear mass effects are appreciable for elements about the middle of the periodic table. Isotope shifts in heavy element spectra (where the mass effect is small or negligible) are discussed in some detail and the theory of the effect of finite distribution of nuclear charge ('volume effect') is outlined. Improved methods of reducing data, including the estimation of isotope shift screening corrections, are referred to, as are various anomalies, including the persistent discrepancy between experiment and calculation, to account for which neutron distribution and nuclear polarization effects have been suggested. A brief description of some magnetic hyperfine structure investigations follows. Small discrepancies in comparisons of nuclear magnetic moments obtained by direct methods, with those from atomic beam magnetic deflection or optical spectroscopic hyperfine structure measurements, have been accounted for in terms of finite distribution of charge and magnetic moment in the nucleus.

§ 1. INTRODUCTION

NUCLEAR properties and effects including the following have been invoked to explain structures found in atomic spectra: (1) the nuclear charge Z, (2) mass M, (3) volume, (4) shape, distribution of charge, electric quadrupole moment Q, (5) effects of neutron distribution, (6) the polarizability of the nucleus by optical electrons, (7) total angular momentum (nuclear spin I) and the associated magnetic dipole moment μ, and (8) distribution of the magnetic moment and effects resulting from the relative motion of nucleons.

Item (1), primarily responsible for determining the atomic energy levels (Meggers 1946, Moore 1949), will not be discussed further. Items (2) to (8)

involve hyperfine structure effects of one sort or another. At the time of the last Report in this series (Crawford 1938), in which the determination of hyperfine structure by optical spectroscopy was considered, the main effort of such investigations was directed to the determination of I, μ and Q for various nuclei. Work of this kind continues—indeed, there has recently been an increasing number of optical measurements. Whilst many such optical determinations are still the only ones available, various radio-frequency techniques have often provided alternative and far more accurate methods of measuring μ and the associated ground (and sometimes metastable) state hyperfine structure. A good deal of the interest in optical hyperfine structure measurements is now turned to the particular type of hyperfine structure known as isotope structure; the only important methods of observing this are those of optical spectroscopy. Although isotope structures (or shifts) have been known for over 20 years, and have been understood in some detail nearly as long, they seem not to have received much general attention; a notable exception is the discussion in the book *Kernmomente*, by H. Kopfermann (1940). In these reports, isotope shifts have only been mentioned, so some of the early developments will be recalled; like I, μ and Q (which have been much more fully considered), isotope shifts, at least in heavy element spectra, are closely connected with nuclear structure.

§2. ISOTOPE STRUCTURE

No doubt all the nuclear characteristics and effects listed in §1, and others besides, affect the hyperfine structure of spectrum lines in some degree. The two main effects, however, are (i) the magnetic hyperfine structure, caused by the interaction of a nucleus possessing spin and magnetic moment with the magnetic field set up at the nucleus by the orbital electrons, and (ii) the isotope structure. The two effects often occur together unless (as is very desirable, though seldom yet achieved) observations (Koch and Rasmussen 1949, Arroe and Mack 1950) can be made with separated isotopes. The components of isotopes of even mass number usually appear single (in these cases it is generally assumed that both I and μ are zero), their wave number differences being the isotope shifts for the transition in question. On the other hand, each of the odd isotopes (which have non-zero I and μ) may show several magnetic hyperfine components for the same transition. The position of the 'centroid', or the equivalent component, of an odd isotope pattern can be determined most accurately by making a magnetic hyperfine term analysis. In this, the position of a term centroid is obtained in the usual way from empirically determined positions and the statistical weights $(2F+1)$ of the sublevels, each characterized by the hyperfine quantum number F. The wave number of the equivalent odd isotope line is then the difference between the centroids for the upper and lower terms. An isotope shift can then be assigned between a single even isotope component and the several components of an odd isotope. From this an isotope shift term analysis can be carried out for all the isotopes (§4(ii)).

Two main kinds of isotopic structure are recognized, usually known as the nuclear mass effect and the nuclear volume effect. The former is believed to be associated mainly with the spectra of light elements and the latter with heavy element spectra. There is no empirical method of deciding how much of each is present in a given case, so that calculations have to be made, or one or the other effect is ignored according as the element is heavy or light.

§3. THE NUCLEAR MASS EFFECT

The simplest examples of nuclear mass effect are those discussed in all textbooks of spectroscopy for hydrogen-like spectra; for example, the displacement of H_α from D_α,* the explanation of which dates back to the time of the Bohr theory in connection with the variation of the Rydberg constant in the spectra of H and ionized He, etc. This variation originates in the fact that the mass of the electron is not negligible compared with the mass of the nucleus, and it will be recalled that the effect was explained by assuming, instead of an electron revolving about a stationary nucleus, that both the nucleus and the electron revolved about their common centre of mass, the two orbits changing simultaneously in a transition.

The same isotope shift formula results from the quantum mechanical treatment of the hydrogen-like atom when all interactions except the electrostatic force between the nucleus and the electron are neglected. The observed displacements for H_α and D_α agree with the theory within the experimental error (which is rather large for such light elements on account of the Doppler breadth of the lines).

For consideration of isotope shifts from this point of view it is convenient to eliminate the Rydberg constant and the principal quantum number and to introduce the term value instead. The displacement ΔT of a term corresponding to an increase of nuclear mass from M to $M+\delta M$ then assumes the form

$$\Delta T=(\Delta\mu/\boldsymbol{m})T \qquad \ldots\ldots(1)$$

where $\boldsymbol{m}$ is the electronic mass, T is the absolute term value (i.e. relative to the series limit) and $\mu=\boldsymbol{m}M/(M+\boldsymbol{m})$ is the 'reduced mass' of the electron for the isotope of mass M. The displacement observed in a spectrum line is then the difference $\Delta T_1-\Delta T_2$ for the terms T_1 and T_2 involved in the transition. For a nucleus of finite mass there is a general contraction of the energy level system relative to that for a nucleus of infinite mass. Consequently, the component of the heaviest isotope has the highest frequency and the terms of this isotope lie deepest relative to the series limit, where $\Delta T=0$ according to (1).

In atoms with more than one electron, since the optical electron interacts in some degree with all the rest as well as with the nucleus, the problem is much more complex. It was first formulated by Hughes and Eckart (1930) who found that, so long as the spin–orbit (and smaller) interactions of the optical electron could be omitted from the potential function used in evaluating the effect, the simple Bohr mass effect equivalent to equation (1) could be separated out in the theoretical treatment. This they called the *normal effect*. The remainder of the calculated isotope shift characteristic of the many-body problem was termed the *specific effect*. The calculations are outlined in several of the papers quoted below and also in Bethe and Bacher (1936) and in Condon and Shortley (1935).

Hughes and Eckart (1930) and Hughes (1931) solved the specific mass effect problem for the isotopes ^{6}Li, ^{7}Li in the I and II spectra and found satisfactory agreement with available data (see also Jackson and Kuhn 1939). Similar calculations have been made for Ne I by Bartlett and Gibbons (1933), for B I by Opechowski and deVries (1939), for Mg I by Vinti (1939), and for He I using the Hughes and Eckart formulae, by Bradley and Kuhn (1948), Andrew and Carter (1948), and Fred, Tomkins and Brody (1949, 1950). For Mg I where there are

* In this particular case the multiplet or fine structure splitting is smaller than the isotope shifts, whilst the hyperfine structure has not been resolved by optical methods.

extensive and accurate data due to Mundie and Meissner (1944) (and several others) obtained by strong electronic excitation of a Mg atom beam, and also for Ne I (Nagaoka and Mishima 1930, 1934, Ritschl and Schober 1937) the agreement is poor. The centroid of ^{25}Mg, however, lies closely halfway between the ^{24}Mg and ^{26}Mg components, as it should according to the mass effect theory.

Calculations of the specific effect are difficult, and very accurate wave functions are needed to obtain a reliable result. Thus, for Ne and Mg it is not yet possible to say definitely whether the discrepancy is due to inaccurate wave functions and approximations of the calculation, or whether there is an appreciable part of the isotope shift which cannot be accounted for as a mass effect. For non-hydrogen-like spectra one cannot say, without a detailed calculation, whether the heaviest or the lightest isotope terms should lie deepest.

Fred, Tomkins and Brody (1949) have tested the Hughes–Eckart theory for the comparatively simple case of He I, for which the required wave functions are more accurately known than for other elements. They used a liquid nitrogen cooled source excited by an electrodeless discharge in a mixture of 55% ^{3}He, 45% ^{4}He at a total pressure of 5 mm. Hg. Some typical results are quoted in Table 1 to illustrate the general agreement with the theory; these (except for $n=2$ terms) and the other results are plotted in Figure 1 to illustrate the difference between theory and observation.

The assignment of shifts to terms was made by calculating the total (normal and specific) shift for the $2\,^1$P and $2\,^3$P terms, where it was assumed correct. The 'observed' shifts in the other terms in the singlet and triplet systems were deduced from the combinations with these two terms (there are no strong intercombinations). The ΔT_{calc} of Table 1 represent the total calculated shift for the P terms; for the S and D terms the ΔT_{calc} are the ΔT of equation (1) S and D terms having no specific effect in the Hughes–Eckart approximation. The differences (column 5) of the P terms, not included in Table 1, seem quite small.

It appears that the fine structure of the $2\,^3$P term is about 15% larger in ^{3}He than in ^{4}He, a matter being further investigated. This may be connected with the discrepancies of Figure 1 (experimental errors are estimated at $\pm 0{\cdot}005\,\text{cm}^{-1}$, whilst the error of calculation resulting from uncertainty of the wave functions of the 2 P terms appears to be of the order of $0{\cdot}01\,\text{cm}^{-1}$). From a later interim report (Fred, Tomkins and Brody 1950), in which some of the ^{3}He magnetic hyperfine structure is resolved, and the effect allowed for in the computation of the isotope shifts, it appears that appreciable discrepancies between observation and the Hughes–Eckart approximation still persist. There is evidence of a perturbation between ^{1}D and ^{3}D terms, especially at $n=8$. Even so, these results probably constitute the best test so far of the Hughes–Eckart approximation. However, it may be that the spin–orbit coupling effect will also need to be included. In other calculations of specific mass effect, referred to above, this interaction has been neglected.

The normal mass effect becomes negligible towards the end of the periodic table. It is often assumed that this also applies to the specific effect, since for a given mass difference between the isotopes, the specific, like the normal shift, varies inversely as the square of the mass, and available calculated specific mass shifts for the light elements happen to show a tendency to decrease with increasing mass and to be of the same order of magnitude as the normal effects. However,

the specific effect depends also on the number of electrons in the atom and contains factors which may increase with the mass in some cases. Bartlett and Gibbons (1933) remarked that where complex configurations are involved, like the $4d^9 5s^2\,^2D^*$ terms in Cu I, the absence of one electron from the 4d shell is likely to have a much bigger specific mass effect than the presence of two s-electrons. However, so far as is known, isotope shifts tend to become small in the middle of the periodic table compared with the magnitudes which they assume at both

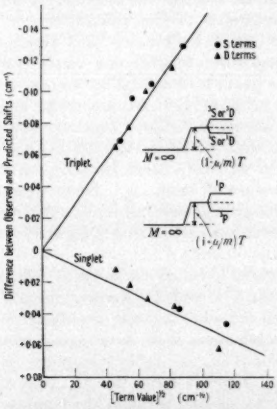


Figure 1. Difference between observed and predicted ^{3}He ^{4}He isotope shifts ($n \geq 3$) plotted against the square root of term value. (*After* Fred, Tomkins and Brody 1949)

Table 1. Isotope Shifts of He Term Values (cm^{-1})

Term	T	$-\Delta T_{obs}$	$-\Delta T_{calc}$	Difference
$2\,^3S$	38,455	1·747	1·758	+0·011
4	8,013	0·494	0·366	−0·128
7	2,242	0·181	0·112	−0·069
$2\,^1S$	32,033	1·534	1·464	−0·070
4	7,371	0·300	0·337	+0·037
$2\,^3P$	29,224	(0·772)	0·772	0·000*
$2\,^1P$	27,176	(1·614)	1·614	0·000*
$4\,^1D$	6,864	0·278	0·314	+0·036
$4\,^3D$	6,866	0·429	0·314	−0·115

* Assumed (figures for several other terms in each series are omitted)

ends. It is certain that the specific mass effect alone is incapable of explaining the isotope shifts found in the heavy elements, for reasons given in §4(i). It is also found that a consistent (though incomplete) explanation of some of the heavy element shifts can be obtained by neglecting the mass effect altogether. There seems little doubt (see end of §4(iv)) that for these heavy elements the mass effect is small (and possibly negligible). An attempt has been made by Dickinson (1934) to account for isotope shifts in Pb I as specific effects, apparently with some

success, but Bethe and Bacher (1936) have raised objections to the calculations on the ground that LS-coupling was assumed, whereas it is, in fact, more like jj-coupling. Presumably, therefore, this could account for Dickinson's results.

From investigations of the Zn I resonance lines, Table 2, Crawford, Gray, Kelly and Schawlow (1950) conclude that for intermediate elements the specific mass effect is by no means negligible. An atomic beam source was used.

Table 2. Isotope Shifts in Zn I Resonance Lines

Transition λ	Observed shift ^{64}Zn,^{68}Zn ΔS	Normal mass effect shift ΔT (eqn. (1))	Residual shift $\Delta S - \Delta T$
$4s^2\ {}^1S_0 - 4s4p\ {}^1P_1$ 2139	$+0{\cdot}033 \pm 0{\cdot}002$	$+0{\cdot}023$	$+0{\cdot}010$
$4s^2\ {}^1S_0 - 4s4p\ {}^3P_1$ 3076	$+0{\cdot}046 \pm 0{\cdot}001$	$+0{\cdot}016$	$+0{\cdot}030$

The shifts between ^{64}Zn, ^{66}Zn are half of the above. The + sign means that the component of the heaviest isotope has the highest frequency. The volume effect theory (§4(iii)) predicts almost identical shifts for these two lines of about $-0{\cdot}03\ \text{cm}^{-1}$. The residual shifts cannot therefore be due to this cause. The difference between the residual shifts, $0{\cdot}020 \pm 0{\cdot}003$, corresponds to a single integral of the mass effect theory. This integral (quoted, but not evaluated, by the authors) has the sign of the residual shift difference, which must accordingly be regarded as a specific mass effect. Moreover, since the average of the residual shifts is $+0{\cdot}020\ \text{cm}^{-1}$ and the volume effect is about $-0{\cdot}03\ \text{cm}^{-1}$ (which is probably an upper limit), the average specific mass effect (the expression for which is also quoted) must be about $+0{\cdot}05\ \text{cm}^{-1}$ (the observed residual shifts being the resultant of the two effects, assuming there are no shifts of unknown origin present). It is also indicated that in view of this large specific effect, some anomalous shifts for Cu I and Zn II discussed by Schüler and Westmeyer (1933) can now be explained qualitatively. Thus for elements of intermediate mass, shifts not due to the mass effect cannot generally be obtained without first finding the specific mass effect by numerical calculation. The nature of the integrals which have to be evaluated to get these specific mass shifts is such that precise values can be obtained only when very accurate wave functions are used. Until accurate integrals are available, it is stated, the prospect of getting significant information about nuclei of intermediate mass from the magnitude of the isotope shifts is not good (cf. discussion of Figure 2, §4(iv)).

§4. THE NUCLEAR FIELD EFFECT

(i) *Nature of the Observational Material*

In the spectra of W I, Pt I, Hg I, II, III, Pb I, II, III, IV, and usually at this end of the periodic table, it has been observed that the various even isotope components and odd isotope centroids in a spectrum line are not linearly spaced with mass as required merely as a consequence of the mass effect; this cannot therefore be solely responsible for the shifts (see also end of §4(iv)). In the elements listed, and in most others examined, the even isotopes happen to be equally spaced to within about 10%, but the centroids of the odd isotopes are 'staggered', falling distinctly away from the midpoint of the adjacent even isotopes, towards the lighter of these. In Sm and Nd there is no approximate

equality of even isotope spacing, but between ^{148}Nd, ^{150}Nd and ^{150}Sm, ^{152}Sm a jump occurs of nearly double the separations between other even isotope components different by two mass units. In all these cases the spacing of the components (or centroids) is in order of mass number. Arroe (1950) now reports that in Ba, where the isotope shifts differ from the above in that they are comparatively very small, the components and centroids are not even in order of mass number. In the Ba I resonance line $6s^2\,{}^1S_0$–$6s\,6p\,{}^1P_1$ the order is 138, 136, 134, with the centroids of 137 and 135 following. In the Ba II resonance lines $6s^2\,{}^2S_{1/2}$–$6p^2P_{1/2,\,3/2}$ the order is 138, 136, 137, 134 and 135. In any one of these three Ba lines the spacing between adjacent pairs of even isotopes is about the same and about equal to the spacing between the centroids of the odd isotopes.

Numerous publications concerning hyperfine structure in Hg, Tl and Pb spectra, especially by Schüler, Keyston and Jones in 1931 and 1932, confirmed that the largest isotope shifts in these elements usually occur in transitions between configurations containing different numbers of s-electrons, especially the deeply penetrating 6s electrons. Transitions of the type $5d^n\,6p$–$5d^n\,6d$ (not involving a change in the number of s-electons) tend to have small or unobservable shifts, whereas much larger and approximately equal shifts, in any particular spectrum, are usually found for such transitions as $5d^n\,6p$–$5d^n\,6s$ and $5d^{n-1}\,6s\,6p$–$5d^{n-1}\,6s^2$. Shifts of about twice this magnitude usually occur for $5d^n\,6p$–$5d^{n-1}\,6s^2$, for example, where there is a change of two in the number of s-electrons. (In W I, Pt I, Hg II, Pb IV, for example, $n=4$, 8, 9, 9 respectively.) The regularities mentioned are sometimes obscured by inter-configuration perturbations and screening effects (§4(v), (vi)).

It will have been gathered that, except for Ba, the *relative* spacings of the various components (or centroids)* are more or less constant and characteristic of the element concerned rather than of the particular lines involved (whether or not these belong to the spectrum of the neutral atom or to one of its ions). This is true to a good approximation for even isotopes where variations in relative spacing outside the experimental error have seldom been observed. Very little is known of the nature or origin (§4(viii)) of the phenomenon. On the other hand, a few relative spacing variations of odd isotope centroids are well established; an extreme example of such variation, unexplained at the time of writing, occurs in the case of Ba just mentioned; another example which has been discussed quantitatively occurs in Hg I (§4(vi)). A nuclear electric quadrupole moment could also cause a relative shift variation for odd isotopes. Also the same causes that produce the even isotope variations may operate for odd isotopes.

The best examples of probable relative variations of even isotope spacing seem still to be the observations of Schüler and Keyston (1931) on the Hg I lines 6072 and 6716 (in 6072 thirteen of the predicted fifteen components were resolved, whilst for 6716 the corresponding figures are probably six out of nine). For the even isotopes, the figures of Table 3 for the separations of components were reported.

The common "$8\,{}^1P$" term may be partly $5d^{10}\,6s\,8p$ in origin, but it is very strongly perturbed from $5d^9\,6s^2\,6p$. If one *arbitrarily* (§4(iii)) assigned zero separations of the isotope levels in the "$8\,{}^1P$" term the figures for 6072 represent

* By the term relative spacings is meant here the ratios of the spacings, e.g. in Hg these would be (^{198}Hg–^{200}Hg)/(^{200}Hg–^{202}Hg) or (^{198}Hg–^{199}Hg)/(^{200}Hg–^{202}Hg) etc.

the shifts in the $^3S_1$ term relative to this datum, whilst those for 6716 correspond for $^1S_0$. Evidence of a similar effect amongst the $5d^9 6p - 5d^8 6s^2$ transitions in Hg III appears from the work of Foster (unpublished) together with the results of Mrozowski (1942) and Foster (1950). On the other hand, Murakawa (1950) finds that within his (unstated) experimental error the even isotope spacings in several Hg I lines, including 6072 and 6716, are constant, but it is highly questionable whether he attained the resolution of Schüler and Keyston.

Table 3. Isotope Shifts in λλ6072 and 6716 Hg I (cm^{-1})

Isotope	198	200	202	204
$6072 = 5d^{10}6s\,7s\ ^3S_1 - "8\ ^1P_1"$	0·115	0·118	0·122	
$6716 = 5d^{10}6s\,7s\ ^1S_0 - "8\ ^1P_1"$	0·121	0·120	0·118	

Manning, Anderson and Watson (1950), using Pb enriched in ^{204}Pb, apparently find no significant variations in the relative spacings of ^{204}Pb, ^{206}Pb, ^{208}Pb components in a number of transitions in Pb I between the $5d^{10}6s^2 6p^2$ configuration and terms of $5d^{10}6s^2 6p\,7s$, 6d, 7d (see also Geiger 1950). Koch and Rasmussen (1950), however, find evidence of relative shift variations between the ^{132}Xe, ^{134}Xe, ^{136}Xe isotopes in different lines. In general, observations with separated isotopes, whereby the accuracy could often be increased ten- or one hundred-fold, are needed to establish the nature of this phenomenon, which may only be detectable at present in rather rare cases.

(ii) *Assignment of Isotope Shifts to Spectrum Terms*

Since there are no combinations between terms of different isotopes of an element one cannot empirically assign absolute isotope shifts to terms (except perhaps for hydrogen). An empirical assignment of isotope displacements is easily made by supposing, arbitrarily, that the isotope levels coincide for any chosen term, whence by application of the combination principle to the observed shifts in various lines connecting through to this term, shifts relative to the chosen datum are assignable directly in other terms. Assignments of this kind show that for terms with high quantum number (but not necessarily for all highly excited terms of the spectrum) the isotope shifts tend to become equal. This is clear otherwise from the fact that the majority of lines in a spectrum, which originate in transitions between high series members, have very small or unresolvable shifts.

A datum of zero shift at the series limit of the spectrum conforms with the above and with simple theoretical considerations, since for any term in which the optical electron has a large principal quantum number, the difference in the binding energy of this electron to the nucleus in different isotopes (whatever the cause) would be expected to be small and to tend to zero at the series limit. If term shifts are assigned relative to this datum they will approximate to the shifts due to the optical electron, the more so the smaller the interaction between the optical electron and other electrons in the atom.

One of the main aims of isotope shift investigations in the heavy elements, with which we are here concerned, is to deduce the shifts associated with individual electrons (usually the 6s electron) from observed term shifts. Thus the importance, firstly of referring the shifts to the series limit, and secondly, of estimating the effect of the interaction of the optical electron with the rest of the atom, is evident. The series limit datum itself at once raised difficulties since

lines actually from the limit are unobservable. Thus an extrapolation from the shifts of observable lines of a series should be made. To add to the difficulty, it not infrequently happens that some of the terms of a series are perturbed by a term of the 'complex' type, which does not belong to the ordinary series spectrum. Such complex terms may lie high in the term scheme, and there may be mutual perturbation between these and certain of the series terms. The complex terms usually have different isotope shift properties from the normal terms which they perturb. The perturbing terms share their properties, including their isotope shifts, so that in such a case the series terms acquire abnormal shifts and extrapolation becomes more difficult. The manner in which corrections to the observed term shifts have been applied to yield the corresponding electronic shifts will be apparent from the subsequent sections of § 4.

Although for purposes of numerical assignment of shifts to terms one can, as has been usual, call the term shift zero at the series limit, any such choice is arbitrary—though there is nothing arbitrary about the shift due to the optical electron being zero at the series limit. Suppose that the spectrum of the neutral atom has been under consideration; its series limit is the ground state of the succeeding ion. For this ion a datum for term shifts is the series limit of the ion, with respect to which the ground state of the ion will have an isotope shift which can be called $S1$. It would be just as good (and in some cases better) to use $S1$ as the datum of the neutral atom term shifts; in this case all the previous values would be increased by the quantity $S1$ (to be subtracted from the term shifts again to approximate to the optical electron shift). One could imagine this process carried on for successive ionizations until all the electrons in the atom were removed, when each of the term shifts in the neutral atom would be increased by $S1+S2+\ldots .S(z-1)$. The shifts of the ground state of the neutral atom would then represent the total isotope shift of the atom. These would be the differences in the energy required to remove all the electrons from the atom in the ground state for each of the various isotopes. For heavy elements the result would probably amount to tens of thousands of wave numbers (cf. Table 6) or several electron volts.

If the datum is zero shift in an arbitrary term the lighter isotope lies deeper in a term with more 6s electrons, as recognized by Breit (1932) for Hg, Tl and Pb spectra. This holds generally for the very heavy elements. Thus, in the ground state of Hg II $5d^{10}6s$, the lighter isotope lies deepest with respect to zero shift at the series limit $5d^{10}$, i.e. the 6s electron is more tightly bound to the nucleus for the lightest isotope; similarly for other terms of the spectrum showing isotope shift. But in Pb II (ground state $5d^{10}6s^26p$) relative to the same datum at the series limit ($5d^{10}6s^2$), the $5d^{10}6s6p^2$ isotope levels are 'inverted'; they are also perturbed. In a case of this kind a 'normal' level scheme can be obtained by supposing that the shifts are referred to the series limit of a higher ion—in this case Pb IV with the series limit $5d^{10}$ containing no 6s electrons.

(iii) *Theory of Isotope Shift in Heavy Element Spectra* ('*Volume Effect*')

A large number of observations of types just described can be co-ordinated by what has been termed 'Breit's theory', the 'nuclear volume effect' and the 'nuclear field effect', by different authors. A remarkable result is that, with only a few exceptions, isotope shifts calculated on the usual assumptions of this theory have been found to be two or three times larger than those observed.

This is not understood quantitatively, and there is some doubt about the qualitative interpretation, but it appears to be not inconsistent with what might be expected from current ideas of nuclear shell structure derived from other sources.

In the theory outlined below it is supposed that isotope shifts are dependent on the finite size of the nucleus and the consequent failure of the Coulomb law at different distances from the centre, corresponding to different radii for different isotopes. The first calculations on these lines were made by Bartlett (1931), Racah (1932), Rosenthal and Breit (1932) and Breit (1932). Until the last few years when further calculations have been made by Broch (1945), Brix and Kopfermann (1949) and Crawford and Schawlow (1949), there appears to have been little further quantitative development of the theory. The relationship of the theory to Schmidt's (1937) theory of the nucleus has been discussed by Fierz (1947), Brix and Kopfermann (1947) and Breit (1950 a, b). Since the time of the early theoretical work notable comparisons of experimental data with the theory have been made by Kopfermann (1940), Brix and Kopfermann (1949), and Crawford and Schawlow (1949).

Different approaches to the theory of isotope shifts have been made by Breit, Arfken and Clendenin (1950) and Ivanenko and Rodichev (1950), in the former instance from considerations of nuclear polarization. An English abstract of the Russian paper states that an "additional exchange effect between nucleons and electrons is postulated".

The isotope shift associated with an s-electron can be calculated for various nuclear models on the following lines. The simplest model is that treated by Racah (1932), in which the nucleus is regarded as a sphere, radius r_0, with the charge Z uniformly distributed over the surface. There is a finite probability of electrons of the outer structure of the atom penetrating into the nucleus. Suppose that the optical electron is considered. The potential energy V of the electron,* when it is inside the nucleus, is then constant and given by $V=-Ze^2/r_0$. For a point nucleus generally, and for the extra-nuclear region, $r>r_0$, of the above model, the potential energy at a distance r is $V_p=-Ze^2/r$ due to the Coulomb field.

The decrease in the binding energy of the optical electron for the case of the finite size nucleus as compared with the point nucleus can be calculated by the perturbation method. The perturbing energy is $V-V_p$, so that the energy change for the whole nuclear region is

$$\Delta W=\int_0^{r_0}\phi^2(r)(V-V_p)4\pi r^2\,dr, \qquad \ldots\ldots(2)$$

where $\phi^2(r)$ is the probability density of the electron at the distance r. According to Racah (1932) for an s-electron,

$$\phi^2(r)=\frac{2(1+\rho)}{[\Gamma(2\rho+1)]^2}\psi^2(0)\left(\frac{2Zr}{a_H}\right)^{2\rho-2}, \qquad \ldots\ldots(3)$$

where $\rho=(1-Z^2\alpha^2)^{1/2}$, $\alpha=2\pi e^2/hc$, the fine structure constant, $\psi^2(0)$ = the square of the non-relativistic wave function at $r=0$ for the state of the s-electron

* Condon (1936) suggested that isotope shifts might be accounted for by a short range electron–neutron interaction. He indicated what magnitude of force would be needed to explain the effect. It has since been shown by Fermi and Marshall (1947) and Rainwater, Rabi and Havens (1949) that electron–neutron forces are small. They appear to be several orders of magnitude smaller than would be needed to explain the isotope shift according to Condon's calculation. Such non-electrical forces are assumed to be absent in this discussion.

concerned, $a_H = \hbar^2/4\pi^2 me^2$, the radius of the first Bohr orbit for hydrogen. Substituting for V and V_p and putting (3) in (2) and integrating,

$$\Delta W = \frac{8\pi(1+\rho)}{[\Gamma(2\rho+1)]^2}\left(\frac{2Z}{a_H}\right)^{2\rho-2}\psi^2(0)\frac{Ze^2r_0^{2\rho}}{2\rho(2\rho+1)}. \quad \ldots\ldots(4)$$

This is the required decrease in the binding energy of the electron, which corresponds to the falling off of the potential from the Coulomb value in the region $r<r_0$. What can be observed experimentally is the variation of this quantity $\delta\Delta W$ from one isotope to another, corresponding to a change in the nuclear radius δr_0. Differentiating (4) with respect to r_0, one obtains

$$\delta\Delta W = \frac{8\pi(1+\rho)}{[\Gamma(2\rho+1)]^2}\left(\frac{2Z}{a_H}\right)^{2\rho-2}\psi^2(0)\frac{Ze^2r_0^{2\rho-1}\delta r_0}{2\rho+1}. \quad \ldots\ldots(5)$$

This is the isotope shift for an s-electron and it is thus relative to the state when the electron is removed from the atom. This shift should be that of the corresponding s-term (referred to zero shift at its series limit), on the assumption that the addition of the s-electron to the ion does not itself affect the isotope levels of the ion (§4(v), (vi)).

If the quantities $y_0 = 2Zr_0/a_H$ and $R = 2\pi^2me^4/ch^3$ (the Rydberg constant) are introduced into (5) the s-term shift (in wave numbers) becomes

$$\delta T = \frac{\delta\Delta W}{hc} = \frac{4\pi Ra_H^3\psi^2(0)}{Z}\frac{(1+\rho)}{[\Gamma(2\rho+1)]^2}y_0^{2\rho}\frac{\Delta y_0}{y_0}B, \quad \ldots\ldots(6)$$

where
$$B = 1/(2\rho+1). \quad \ldots\ldots(7)$$

The internal potential field of the nucleus can be varied without affecting (6), according to the above method of calculation, apart from changing the last term, B. For example, if the charge is supposed uniformly distributed throughout the volume of the nucleus (Breit 1932) $V = [-\frac{3}{2}+\frac{1}{2}(r/r_0)^2]Ze^2/r_0$ and

$$B = 3/(2\rho+1)(2\rho+3). \quad \ldots\ldots(8)$$

If $V = -Ze^2/r_0$, and if it is the same for each isotope, then obviously

$$B = 0. \quad \ldots\ldots(9)$$

One could imagine that this corresponds to an extreme case where the added neutron stays on the outside of the nucleus without altering the proton distribution, which we are assuming determines the shift.

Other spherical charge distributions have been considered by Rosenthal and Breit (1932) and Crawford and Schawlow (1949), whilst Brix and Kopfermann (1949) have dealt with an ellipsoidal distribution.

The formula now used for the non-relativistic $\psi^2(0)$ in (6) is

$$\psi^2(0) = \frac{Z_1Z_0^3}{\pi a_H^3 n_0^3}\left(1-\frac{d\sigma}{dn}\right) \quad \ldots\ldots(10)$$

as developed by Goudsmit (1933) and Fermi and Segrè (1933), where Z_1 is the effective nuclear charge in the inner region of the atom, put equal to Z for an s-electron, Z_0 is the effective nuclear charge in the outer region (for an optical electron it will be 1 for neutral, 2 for singly ionized atoms, and so on), n_0 is the effective quantum number, and σ the quantum defect. If T is the absolute term value these quantities are related as follows:

$$T = RZ_0^2/n_0^2. \quad \ldots\ldots(11)$$

If the terms are unperturbed, they do, and should, as a consequence of quantum mechanics (Shenstone and Russell 1932) fit the Ritz series formula

$$T = RZ_o^2/(n-\alpha-\beta T)^2, \qquad \ldots\ldots(12)$$

where n is the principal quantum number and α and β are constants for a given term series (e.g. $ns^2S_{1/2}$) obtainable from observational data. Finally,

$$\sigma = \alpha + \beta T, \qquad \ldots\ldots(13)$$

and from (12 and 13) follows the result

$$d\sigma/dn = \beta/(\beta - n_0/2T), \qquad \ldots\ldots(14)$$

used by Crawford and Schawlow (1949). Substituting (8), (10) and (11) in (6) the s-term shift in Crawford and Schawlow's form becomes

$$\delta T = \left[\frac{4T^{3/2}}{Z_o R^{1/2}}\left(1 - \frac{d\sigma}{dn}\right)\right]\left[\frac{3}{(2\rho+1)(2\rho+3)}\right]\left[\frac{(1+\rho)y_0^{2\rho}}{[\Gamma(2\rho+1)]^2}\right]\left[\frac{\delta y_0}{y_0}\right], \qquad \ldots\ldots(15)$$

The first term in (15) is the $\psi^2(0)$ expression in (10), apart from a factor proportional to Z. The next two terms, of which the first is B (see equations (7), (8), (9)), which is characteristic of the nuclear model, depend on the relativistic correction to $\psi^2(0)$ resulting from the solution of the Dirac equation for the optical electron in the electrostatic field of a point nucleus. The last term is equal to $\delta r_0/r_0$, the fractional change in the nuclear (i.e. the proton distribution) radius in passing from one isotope to another.

In the early work of Racah (1932) and Rosenthal and Breit (1932) the $\psi^2(0)$ was obtained by numerical solution of the wave equation. Much more accurate values are given by (10), which has been widely used for the reduction of hyperfine structure (Kopfermann 1940) to obtain nuclear magnetic moments. There can be little doubt that (10) is substantially correct (§5).

Calculation of the errors in the above first order perturbation treatment are mentioned in §4(vii).

Rosenthal and Breit (1932) have calculated the shifts for non-s-electrons which, for the same principal quantum number, were found to be negligibly small by comparison except for $p_{1/2}$-electrons. For a $6p_{1/2}$-electron the shift should be of the order one-twentieth that of a 6s-electron.

(iv) *Comparison of Experimental Data with Volume Effect Calculations*

Although a considerable collection of isotope shift data has been published (the references quoted here are by no means complete), there is no great wealth of material with which to make critical comparisons with the volume theory, one reason being that most existing measurements are insufficiently complete for proper 'normalization' (to use Kopfermann's term) of the shifts to the series limit datum. Secondly, much work, usually difficult, remains to be done in the way of detailed applications of spectroscopic theory to particular cases.

At the end of §4(iii) the relatively small values of the calculated shifts for non-s-electrons was remarked. Assuming the truth of this, an approximate series limit normalization may be obtainable by referring the shifts, instead of to the series limit, to a suitable term containing a $p_{3/2}$ or $d_{3/2}$ or $d_{5/2}$ electron (in place of the 6s or 5s electron for which the shift is required). Alternatively, a small shift calculated by the volume theory can be applied to a term with, say, a $p_{1/2}$-electron

or perhaps an s-electron with $n>6$. It is on these lines that most limit normalizations have been attempted.

Crawford and Schawlow (1949) have given the most critical discussion of the $\psi^2(0)$ dependency of isotope shifts by considering the comparatively simple cases of the 'one electron spectra' of Hg II, Tl III and Pb IV. The observed term shift normalized as well as possible to the series limit is equated to δT given by equation (15), whereby an 'observed' apparent value of $\delta y_0/y_0=\delta r_0/r_0$, the fractional change in the nuclear radius, is obtained. It was assumed that

$$r_0=1{\cdot}5\times10^{-13}A^{1/3}\,\text{cm}. \qquad \ldots\ldots(16)$$

where A is the mass number (Bethe 1947). The results are given in Table 4.

Table 4. Apparent Change in Nuclear Radius $\delta y_0/y_0=\delta r_0/r_0$
(for uniform charge distribution throughout the nucleus)

	Hg II		Tl III		Pb IV	
(1)	(2)	(3)	(2)	(3)	(2)	(3)
$5d^{10}$ 6s	0·276	1·05	0·38	1·13	0·5	1·(2)
7s	0·050	1·02	0·090	1·01		
8s	0·080*	4·2	0·046	1·2		
9s	0·10±0·025*	11				
10s	0·05±0·03*	10				

(1) level; (2) isotope shift (cm^{-1}); (3) $(\delta r_0/r_0)10^{-3}$ apparent.

* Perturbed.

The values of $\delta r_0/r_0$ are termed apparent because they have to be corrected for screening effects and the error in the first order perturbation calculation (§4(vi), (vii)). The $n=8$, 9, 10 terms in Hg II are perturbed probably by the (unidentified) $5d^9\,6s\,6d$ configuration, expected about here. The perturbation was demonstrated by non-linearity of the plot of quantum defect σ against term value (equation (13)). Consequently the isotope shifts of these terms are irregular. The Tl terms for which data are given are relatively unaffected, and their $\delta r_0/r_0$ values agree much better. The screening correction for the 6s-electron results in an increase of 16% in the apparent $\delta r_0/r_0$; it is not given for the 7s-electron, but is probably comparatively small, and it is almost negligible for $n\geqslant8$. The consistency of the $\delta r_0/r_0$ in Tl for $n=6$, 7 and 8 is good evidence for the dependence of the isotope shifts on $\psi^2(0)$, the first term of (15). If the volume of the nucleus is proportional to the mass number, and $\delta A=2$, $\delta r_0/r_0=2/3A=3{\cdot}3\times10^{-3}$ for $A\simeq200$. It is seen from Table 4 that $\delta r_0/r_0$ is about 1·1 for Tl and Hg. This value of $\delta r_0/r_0$ for the 6s-electron has to be increased to $1{\cdot}1\times10^{-3}\times1{\cdot}16\times1{\cdot}35=1{\cdot}7\times10^{-3}$ because of the screening and first order calculation errors already mentioned. Thus the 'observed' fractional change in nuclear radius is half that calculated on the basis of nuclear volume proportional to mass and a uniform distribution of charge throughout the volume. This result is substantiated by almost all other observations.

Other comparisons of theory and experiment which, taken together, include a wide range of elements, have been made by Köhler (1939), Kopfermann (1940), Brix and Steudel (1950), Brix and Frank (1950) and especially Brix and Kopfermann (1949). Brix and Steudel (1950) have attempted to eliminate the mass effect in their Cd I results by means of analogy, otherwise the mass effect has been neglected.

Figure 2 illustrates the results; it is substantially Figure 6 in the paper of Brix and Kopfermann. In Figure 2 the experimental and theoretical 'isotope shift constants' C_{th} and C_{exp} are plotted against Z. These are defined as follows:

$$\delta T_{th} = \frac{Z_0^2}{n_0^3}\left(1 - \frac{d\sigma}{dn}\right) C_{th}, \qquad \ldots\ldots(17)$$

where $\delta T_{th} = \delta T$ is given by (15) for r_0 and δr_0 given by (16), δr_0 corresponding to $\delta A = 2$:

$$\delta T_{exp} = \frac{Z_0^2}{n_0^3}\left(1 - \frac{d\sigma}{dn}\right) C_{exp} \qquad \ldots\ldots(18)$$

where δT_{exp} is the experimentally observed isotope shift. The coefficient of C in (17) and (18) differs only by the factor $4R$ from the first term of (15). This coefficient involves only quantities concerning the outer structure of the atom, obtainable if the multiplet analysis of the spectrum is sufficiently complete. C_{th}, however, involves the nuclear model, the relativity effects which occur mainly

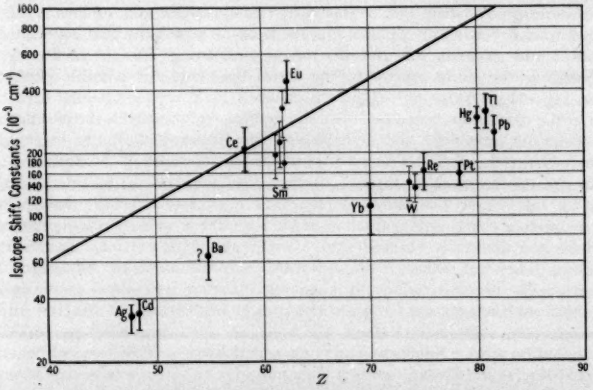


Figure 2. Brix–Kopfermann diagram. Variation of isotope shift constants C_{th} and C_{exp} with Z. —— C_{th} for $r_0 = 1{\cdot}5 \times 10^{-13} A^{1/3}$ with δr_0 corresponding to $\delta A = 2$: ● C_{exp}.

The C_{exp} of Figure 2 were derived from data obtained by the following:

For Ag, Rasmussen (1940 a, b); Cd, Schüler and Westmeyer (1933), Brix and Steudel (1950); Ba, Kopfermann and Wessel (1949); Ce, Brix and Frank (1950); Sm, Brix and Kopfermann (1949); Eu, Schüler and Schmidt (1935); Yb, Schüler and Korching (1938); W, Kopfermann and Meyer (1948); Re, Schüler and Korching (1937); Pt, Jaeckel and Kopfermann (1936), Tolansky and Lee (1937); Hg, Schüler and Keyston (1931), Schüler and Jones (1932 a, c), Schmidt (1938), Mrozowski (1938, 1942); Tl, Köhler (1939); Pb, Schüler and Jones (1932 b).

within and near the nucleus and the nuclear charge. The dependence of δT_{exp} on the coefficient of the C's (which differs by the factor $Z_1/\pi a_H^3$ from $\psi^2(0)$ seems to be fairly well established, for example, from the work of Crawford and Schawlow; as already remarked, the $\psi^2(0)$ expression itself (10) is not in doubt to the accuracy at present required. Thus in considering C_{th} and C_{exp} one should be entirely concerned with nuclear effects. C_{th} depends mainly on Z and not much on the fractional change in the nuclear radius, which only changes from about $\frac{1}{160}$ to $\frac{1}{310}$ in going from Ag to Pb. The Brix–Kopfermann diagram can be

regarded as a test of the non-$\psi^2(0)$ part of equation (15). For most of the elements of Figure 2 C_{exp} corresponds to the average shift per 6 (or 5) s-electron for the most easily observed isotopes of an element differing by two mass units, but where the differences in the behaviour of various isotope pairs are marked, as in Sm, and to a lesser extent in W, the constants for the individual pairs are entered. The tolerances given in the Brix–Kopfermann diagram, for the most part, represent their estimates of the uncertainty in the C_{exp} values. The separations of the Ba I isotope pairs given by Kopfermann and Wessel (1948) are about five times larger than those of Arroe (1950) ($0{\cdot}0022\,cm^{-1}$) which were much better resolved. There appear to be complications here, probably involving perturbation effects, but it would seem that the ordinate for the Ba entry should be much smaller than that corresponding to Kopfermann and Wessel's result in Figure 2.

The most striking feature of Figure 2 is the fact that the C_{exp}, except for Sm and Eu, are consistently smaller than the C_{th} calculated according to (15). This corresponds with the anomalously small 'observed' nuclear radii of Crawford and Schawlow. There appears to be no possibility that experimental errors could account for the discrepancy, allowing for the difficulty of normalizing the shifts exactly to the series limit and for possible effects of screening and perturbation. Thus the cause of the discrepancy must be in the part of the calculation concerned with the nuclear region or with the effect of the particular nuclear model. In the improved calculations, §4 (vii), the discrepancy is somewhat reduced, but a factor of two in the magnitudes of the observed and calculated results persists, so that it is concluded that the nuclear model rather than the calculation is at fault. As Brix and Kopfermann (1949) remark, it is not surprising that improvement is needed on the crude representation of a nucleus as a uniformly charged sphere with volume proportional to the mass number. They had earlier (1947) shown that various models can be constructed which resolve the discrepancy by assuming that the last added neutrons are for the most part disposed outside the proton distribution (if the last added neutrons stay completely outside the nucleus, as a result of equation (9) C_{th} should be zero). Similarly, Crawford and Schawlow (1949) remark that the shifts in Hg, Tl and Pb indicate that the addition of two neutrons to each of these nuclei does not cause a complete redistribution of nuclear charge, a result consistent with the existence of some type of reasonably stable structure in the nucleus such as the shell structures discussed by Feenberg and Hammack (1949), Mayer (1948), Nordheim (1949) and others, which have been contrasted by Feenberg (1950).

The agreement of the observed and calculated isotope shifts is somewhat less satisfactory than that obtained with (15) (which includes equation (8)) if the nuclear model corresponds to equation (7) (uniformly charged spherical shell). But, again, the results can be modified, as above.

The second point of interest in Figure 2 is the sharp increase in C_{exp}, apparently between Ba and Eu. This seems to be genuine enough, in spite of the fact that the rare earth spectra from which results have been extracted are amongst the most complex spectra known and the estimated experimental errors in the determination of C_{exp} ($c.\pm 25\%$) seem somewhat low. Brix and Kopfermann (1949) and Brix and Frank (1950) suspect a connection between this phenomenon and the distinguished neutron number 82, ^{138}Ba having 82 neutrons.

A further correlation between isotope shifts and neutron numbers, Table 5, is made by Brix and Kopfermann (1949) in connection with the *relative* (footnote, p. 294) isotope shifts in Sm (Schüler and Schmidt 1934, Brix and Kopfermann 1949) and Nd (Klinkenberg 1945, who pointed out that anomalously large shifts occurred at the introduction of the 45th pair of neutrons in both Sm and Nd nuclei).

Table 5. Correlation of Relative Isotope Shifts and Neutron Numbers

Sm	Mass number	144		148	150	152	154
	Neutron number	82		86	88	90	92
$Z=62$	Relative isotope shift	0		2	3·14 ±0·06	4·81 ±0·08	5·72 ±0·08
Nd	Mass number	142	144	146	148	150	
	Neutron number	82	84	86	88	90	
$Z=60$	Relative isotope shift	0	1·03	2	3·08	4·73	

For the same neutron numbers *all* the relative shifts agree within the experimental error. If this is not fortuitous it seems that, in spite of the two extra protons in Sm, the relative shifts in both elements are determined by the neutron numbers. Evidence for stable closed shell structures at 82 neutrons was given by Elsasser (1934) and Mayer (1948). The above results can be interpreted as evidence of unclosed neutron shells.

The anomalously large relative shifts between ^{150}Sm, ^{152}Sm and ^{148}Nd, ^{150}Nd have been the subject of special consideration from other points of view. Schüler and Schmidt (1934) pointed out a probable connection of the Sm anomaly with the reported α-activity of Sm. Dempster (1948) showed that ^{152}Sm was the probable emitter. Feather (1948), noting that the abnormal jump in the isotope shift indicates a corresponding abnormal increase in the proton-occupied volume of the nucleus, remarks that this can also be regarded as evidence of a 'discontinuous' decrease in nuclear binding energy as between the pairs of Sm and Nd isotopes concerned. He thus classifies the Sm isotopes 144, 148, 150 and Nd isotopes 142, 144, 146, 148 as tightly bound as regards α-emission and ^{152}Sm, ^{154}Sm and ^{150}Nd as loosely bound. On this basis the only possibility of α-decay from a loosely bound Sm to a tightly bound Nd is $^{152}\mathrm{Sm} \xrightarrow{\alpha} {}^{148}\mathrm{Nd}$, so that, again, there is evidence for the activity of ^{152}Sm. Other evidence also favours this assignment.

A further discussion of the Sm anomaly, based on an α-particle model of the nucleus, has been given by Wefelmeier (1939).

Brix and Kopfermann (1949), however, take another view and associate the anomalous ^{150}Sm, ^{152}Sm shifts with the increase in the nuclear quadrupole moment which occurs between ^{151}Eu and ^{153}Eu. One might perhaps regard the connection as wholly accidental but for the fact that a smaller but similar effect had previously been noted for the ^{184}W and ^{186}W shift and the quadrupole moment change between ^{185}Re and ^{187}Re (Kopfermann and Meyer 1948, Brix and Kopfermann 1947). They assume, in accordance with the Schmidt (1937, 1940) single-particle model of the nucleus (cf. also Feenberg 1950 and Townes, Foley and Low 1949) that ^{151}Eu differs from ^{150}Sm mainly by an unpaired proton, and similarly for ^{153}Eu and ^{152}Sm. The Eu quadrupole moments are regarded as too large to originate in a single proton, so that the 'core' of each of the Eu nuclei 151 and 153 also has a non-spherically symmetric charge distribution, which, apparently, is supposed to be carried over to ^{150}Sm and ^{152}Sm respectively.

For convenience of calculation the charge distribution is regarded as ellipsoidal. The change in the charge distribution from a uniformly charged sphere strained into an equal volume ellipsoid is, from the point of view of the associated isotope shift, equivalent to an increase in volume of the sphere. It is also arbitrarily assumed that the usual increase in total volume occurs for the addition of two neutrons. The supposed greater length of the ^{152}Sm nucleus as compared with ^{150}Sm in accordance with the ^{151}Eu and ^{153}Eu quadrupole moments corresponds with the anomalous shift.* Breit, Arfken and Clendenin (1950) state that this explanation is not literally tenable because isotopes with total angular momentum zero (as assumed for ^{150}Sm ^{152}Sm) must have spherical symmetry, but they agree that the essential idea is applicable and show how a nuclear wave function can be constructed which would allow the effect; but other difficulties arise.

These considerations point to the potentialities of plots of both C_{th} and C_{exp} against Z in general (or somewhat similar plots of $\delta r_0/r_0$ against Z and corresponding plots against neutron numbers) and the relative values of isotope shifts for a given element in particular, as tools for the investigation of nuclear structure. It may be anticipated that as more separated isotopes become available, the experimental data will be very greatly improved and, in addition, interesting results will follow from investigations of radioactive isotopes.

In the heavy element spectra, the mass effect has been neglected. Good evidence that it is small is that the isotope shifts in unperturbed configurations, in accordance with the theory outlined, are characteristic of the number of s-electrons therein rather than of individual terms of the configuration. Terms tending towards singlet and triplet types in any one configuration in many cases would be expected to show larger shift variations than observed, if there was any considerable mass effect. The analysis, however, has not yet been carried to a stage warranting the assumption that mass effects are entirely negligible.

(v) *The Effect of Screening of the Inner Electrons by the Optical Electrons*

For the most part this effect has been neglected in the work described above, but for the precise determination of the shifts associated with individual electrons from the observed term shift the screening effect will need to be considered. The following discussion is based on Crawford and Schawlow (1949).

For the small fraction of the 'orbital period' that the optical electron is closer to the nucleus than any one of the inner electrons of the atom, by a screening action it reduces the effective nuclear charge acting on the inner electron by one unit. By thus reducing the binding energy of the inner electron to the nucleus, it reduces the contribution of the latter to the isotope shift of the whole atom. Thus the mere presence of the optical electron disturbs the series limit datum, and to get the optical electron shift the change in this datum has to be added to the term shift deduced empirically (for example from lines originating from highly excited terms not containing penetrating electrons—it can be assumed that such non-penetrating electrons have a negligibly small screening action on the innermost electrons, compared with penetrating electrons of the 6s type). Since the optical electron only spends a small fraction of time 'inside' the inner electrons, it only causes a small fractional change in their isotope shifts. But the inner electrons are strongly bound to the nucleus and will have large isotope

* A recent recalculation (Sternheimer 1950) considerably increases the Eu quadrupole moments, but a correspondence still remains, qualitatively.

shifts, so that it is not surprising that the change produced is an appreciable fraction of the isotope shift of the optical electron.

If the value of $\psi^2(0)$ were known accurately for the inner electrons when the optical electron is present, and also for the case when it is removed to the limit, the screening effect could be estimated directly. Since the required accuracy in $\psi^2(0)$ is not obtainable directly from self-consistent field calculations, the effect is estimated as follows:

The probability p that a 6s optical electron (for example) is closer to the nucleus than each of the inner s-electrons with $n=1, 2 \ldots . 5$ can be obtained by evaluation of the integral

$$p=\int_0^\infty \left[P_{ns}^2 \int_0^{r_1} P_{6s}^2 \, dr \right] dr_1, \qquad \ldots\ldots(18)$$

P^2 is the probability of the electron being between r and $r+dr$, obtainable if suitable Hartree wave functions are available (Hartree 1946–7); the integral is evaluated numerically.

From equation (6) the isotope shift for an ns-electron and for a 6s-electron should be in the ratio $\psi_{ns}^2(0) : \psi_{6s}^2(0)$. But by equation (10) $\psi^2(0)=Z_1 Z_0/\pi a_H^3 n_0^3$ (neglecting the correcting factor). During the fraction p of the time that the optical electron is closer to the nucleus than the inner electron it reduces Z_0 by 1. For this fraction of the time $\psi^2(0)$ is reduced by the fraction

$$[Z_0^2-(Z_0-1)^2]/Z_0^2 \simeq 2/Z_0.$$

Thus the change in the isotope shift of the inner ns-electron expressed as a fraction of the isotope shift of the 6s-electron is

$$\frac{(\delta T)_{ns}}{(\delta T)_{6s}}=\frac{\psi_{ns}^2(0)}{\psi_{6s}^2(0)}\frac{2}{Z_0}p. \qquad \ldots\ldots(19)$$

The ratios of the $\psi^2(0)$ in (19) are obtainable approximately if the Hartree functions are available, whilst from the shell model of the atom Z_0 is about equal to Z for 1s-electrons, about $Z-2$ for 2s-electrons, $Z-10$, $Z-28$ and $Z-60$ respectively for 3s, 4s and 5s electrons in the heavy elements.

The assumption that n_0 is independent of the screening by the optical electron is valid for the innermost of the inner electrons, but it is only roughly true for the others. The calculated screening corrections are consequently to be regarded as upper limits. The optical electrons also screen the inner p, d and f electrons, but since these should have only very small isotope shifts, the screening correction should be negligible.

Crawford and Schawlow find the results of Table 6, evaluated from the Hartree functions (Hartree and Hartree 1934) for mercury.

Table 6. Isotope Shift Screening Corrections

Electron	$\psi^2(0)\pi a_H^3$	p	Screening correction*
1s	$5{\cdot}06\times10^5$	$2{\cdot}034\times10^{-5}$	0·016
2s	$5{\cdot}56\times10^4$	$1{\cdot}912\times10^{-4}$	0·017
3s	$1{\cdot}23\times10^4$	$9{\cdot}010\times10^{-4}$	0·020
4s	$3{\cdot}06\times10^3$	$3{\cdot}846\times10^{-3}$	0·028
5s	$6{\cdot}10\times10^2$	$2{\cdot}090\times10^{-2}$	0·079
6s	$3{\cdot}2\times10^1$		
		Total	0·16

* For two ns-electrons as a fraction of $(\delta T)_{6s}$.

Thus the shift for the 6s-term of Hg II, for example, has to be increased by the 16% mentioned in § 4 (iv) to give the shift for the 6s-electron.

Crawford and Schawlow re-analyse the extensive Tl II isotope shifts of Köhler (1939), and with the screening corrections taken into account these are shown to be consistent with Tl III data of Convey (1940).

(vi) *Configuration Perturbation*

There are several examples of irregularities in isotope shift phenomena which have been ascribed to interconfiguration perturbation effects of the type which arise between terms of the same parity and J values. A particularly good example (Jones 1933) occurs in the Pb II spectrum (ground state $5d^{10}6s^26p$) between the configurations $5d^{10}6s^26d$ and $5d^{10}6s6p^2$. If the shifts are normalized to the ionization limit of Pb IV (end of § 4 (ii)) one would expect to find shifts about twice as big in $5d^{10}6s^26d$ as in $5d^{10}6s6p^2$ since the former contains two and the latter only one 6s-electron. Actually, the $5d^{10}6s^26d$ shifts are smaller and those of $5d^{10}6s6p^2$ larger than expected. This is to be regarded from the point of view of the mixing of the wave functions for $6s^26d$ and $6s6p^2$, whereby some of the large $\psi^2(0)$ associated with the terms containing $6s^2$ is transferred to the term with 6s. In addition, there will be differential screening effects to be considered. Other examples are to be found amongst Pt I shifts (Jaeckel and Kopfermann 1936, Tolansky and Lee 1937). Here irregularities of the magnetic hyperfine structure (Schmidt 1936) and of the multiplet structure show that some terms of $5d^86s^2$ interact with some of those of $5d^96s$, so that the shifts of some terms of the first-mentioned configuration are less and some of the latter larger than expected. There appears to have been no detailed quantitative consideration of isotope shifts (and very few of magnetic hyperfine structures) from this point of view, apart from that in a paper by Hume and Crawford (1950, abstract only) which deals both with the magnetic hyperfine structure and isotope shifts in Pb III. This abstract states that precise configuration assignments were made for many Pb III terms for which the wave functions were found to be linear combinations of LS functions of two or more configurations. Using these wave functions, it is reported that excellent agreement with the observed g-values, magnetic hyperfine interval factors and isotope shifts was obtained. These isotope shifts were accounted for on the basis of the 'field theory' alone (this probably refers to the volume theory, § 4 (iv)).

More is known, from the work of Breit and Wills (1933) and of Crawford (1935), about the closely related problem of interconfiguration perturbation and intermediate coupling effects in connection with magnetic hyperfine structure.

A special type of magnetic hyperfine structure perturbation (rarely encountered), which, incidentally, can produce term to term variations in the relative spacing of *odd* isotope centroids has been discussed by Crawford (1936) and in more detail by Goudsmit and Bacher (1933) (Hg, Al) and McLay and Crawford (1933) (Bi).

(vii) *Second Approximation in the Calculation of the Volume Effect*

The calculation of the volume effect in § 4 (iii) assumes that the wave function of the optical electron in the nuclear region is the same in the presence of the nucleus of finite size as it is for the point nucleus. The perturbation takes place only over a very small region of space, but the electric fields causing the energy

change are very large, and the effect of the resulting distortion of the wave function has to be investigated. The matter was considered by Rosenthal and Breit (1932), who concluded that the error involved in the first order calculation could not be large enough to bring the observed and calculated shifts which they considered into agreement.

Crawford and Schawlow (1949), using a method due to Broch (1945), estimate that for the spherical shell model of the nucleus (equations (6) and (7)) and for the uniformly charged sphere model (equations (6) and (8)) for an s-electron of Tl, the first order perturbation calculation is in error by 36% and 34% respectively. This error causes part of the necessary increase in $\delta r_0/r_0$ mentioned in the discussion of Table 4, § 4 (iv).

(viii) *The 'Staggering' of the Odd Isotope Centroids*

The 'staggering' of odd isotope centroids (§ 4 (i)) appears to be qualitatively consistent with what would be expected from the Schmidt (1937, 1940) model of the nucleus, as pointed out by Brix and Kopfermann (1947), Fierz (1947) and Breit (1950 a, b). Schmidt showed that nuclear moments could be accounted for approximately using a single particle plus core model of the nucleus. According to this, in first approximation, the mechanical and magnetic moments are supposed to arise from a single nucleon, which preserves the spin and magnetic moment of the free state. The orbital angular momentum of an odd proton is used as though the proton was on the outside of the nucleus, whereby it is assumed to be largely independent of the motion of nuclear particles in the core. An odd neutron is also regarded as mostly outside the core. When the nuclear magnetic moments are calculated on this basis according to Landé's (1933) vector coupling scheme it is found that all observed magnetic moments lie within predicted limits; there thus appears to be substantial evidence that this viewpoint is partly correct. When two neutrons are added to a nucleus with even Z, it often happens that the nuclear mechanical moments are unchanged and the magnetic moments are unchanged or almost so. In this case, according to Schmidt's model, the pair can be considered as forming part of the core (although they need not necessarily wholly redistribute the charge). Since the odd neutron in this model remains mostly outside, and a pair of neutrons becomes considerably more a part of the core, one can account qualitatively for the staggering.

Breit, Arfken and Clendenin (1950) have made elaborate calculations to determine the order of magnitude of polarization effects of a nucleus by the atomic electrons. They find that it is conceivable that large enough effects could arise to account for the staggering of the odd isotopes of Pb, and it is also possible that polarization may be responsible, at least for some of the discrepancy between the observed shifts and those calculated according to the volume theory. The types of experimental data in various fields which would be useful for testing the theory are indicated. The experimental basis of the calculation is largely evidence, derived from observations of radioactive decay, that the low nuclear energy levels are generally more closely spaced in odd than in even nuclei. From this it appears that an atomic electron should cause a greater polarization effect in odd than in even nuclei, whereby the binding energy of the electron to the nucleus is decreased more for the former than for the latter. It is not certain, however, to what extend some of the assumptions made in the calculations are justified (Breit 1950 a), and there appears to be some doubt as to how significant

the effect will be in practice. The polarization theory may also be capable of accounting for some of the small term-to-term variations of relative isotope shift (§4(ii)).

§5. NUCLEAR MOMENTS

A very valuable collection of data on nuclear spin, magnetic and electric quadrupole moments available up to 1st January 1950 has been given by Mack (1950). A bibliography is included.

Sternheimer (1950) has recalculated several nuclear electric quadrupole moments taking into account the quadrupole moment induced in the electron shells by the nuclear quadrupole moment. This has resulted in considerable changes in some previously accepted values which have included the effect of the induced quadrupole moment.

In this section no attempt is made to bring Mack's table up to date or to do justice to experimental work. The emphasis is rather on the significance at present attached to small discrepancies in results from different types of measurement of nuclear magnetic moment.

Optical determinations of nuclear magnetic moments have continued on general lines indicated in earlier Reports (Crawford, 1938, where the earlier reports are quoted). Especially notable have been: (i) the increasing use of atomic beam light sources excited by intense currents of slow electrons (Mundie and Meissner 1944), and (ii) investigations of small samples of rare isotopes separated (or enriched) by mass spectrographic methods (cf. Arroe and Mack 1950).

Direct determinations of μ by nuclear induction (nuclear magnetic resonance, nuclear paramagnetic measurements, Rollin 1948) are rapidly replacing optical measurements. In one form of these experiments the substance investigated may be used in the solid, liquid or gaseous state, the nucleus in question being contained in a diamagnetic atom, molecule or ion, whereby the resultant electronic angular momentum J of the particle is zero. A nuclear Zeeman effect is produced when an external magnetic field is applied. The magnetic field, apart from a slight reduction (usually) by diamagnetic screening of the nucleus by the atomic electrons, is essentially the externally applied field. If the nuclear magnetic moment is μ' and the magnetic field at the nucleus is H, the energy difference between adjacent Zeeman states is $\mu' H/I$. This is $eh\mu H/4\pi MeI = h\nu$ if μ is expressed in nuclear magnetons, M being the proton mass and ν the frequency of transitions which can be induced between adjacent states. Thus if ν is measured and H is known, μ is determined if the spin I is known.

Applications of radio-frequency techniques to atomic beams include two classes of experiment for determination of μ. When the nucleus is in the form of a diamagnetic atom or molecule, a very accurate direct measurement can be made by the method of Rabi, Millman, Kusch and Zacharias (1939) (see also Crawford 1938). This is somewhat analogous and of comparable accuracy to the method referred to in the last paragraph (of which the atomic beam method was the forerunner).

A second type of atomic beam magnetic resonance experiment employs a paramagnetic atom (Kusch, Millman and Rabi 1940, Davis, Nagle and Zacharias 1949). J is no longer zero, and consequently a very intense magnetic field exists at the nucleus due to the atomic electrons. In this case the method

depends on the observation of the Zeeman effect of the whole atom in an externally applied magnetic field (in most experiments the atom has been in the ground state, though some metastable states have been investigated). When the external field is sufficiently weak the coupling between J and I is maintained. The combined system has the maximum total angular momentum $F_{max}=I+J$, but the magnetic moment associated with F is always almost entirely electronic in origin. The magnitude of the Zeeman splitting is now $eH/4\pi mcF=1{\cdot}400\,H/F$ Mc/s.* Sometimes the hyperfine structure $\Delta\nu$ can be observed directly by inducing $\Delta F=\pm 1$ transitions in a weak field. As the magnetic field is increased, the coupling between I and J breaks down and the Paschen–Back effect develops. This involves the interaction of the nuclear moment with the magnetic field of the atomic electrons, in other words the hyperfine structure. If I is known, the hyperfine structure can be calculated from the variation of the resonance frequencies with magnetic field in the stage where the Paschen–Back effect has developed. Having obtained the hyperfine structure, the nuclear moment is obtainable from it by formulae used in reducing the same data obtained by optical investigations. Although the hyperfine structure can be measured much more accurately by this than by optical methods, owing to uncertainty in the formulae representing the interaction between the optical electrons and the nuclear moment, the accuracy of the resulting μ is usually of the same order as is obtainable from the best optical spectroscopic measurements.

An important generalization from the available results of accurate measurements of μ, by direct and hyperfine structure methods is that both methods give substantially the same values. This shows that the usual theory of hyperfine structure separations (cf. Kopfermann 1940) is certainly not far wrong. It also follows that the $\psi^2(0)$ formula (equation (10)) which is involved in the hyperfine structure reduction is substantially correct. In some cases, however, the two types of measurement do give results differing by small amounts definitely outside the experimental error. Some of these discrepancies have been explained in terms of effects resulting from the finite size of the nucleus which affect the hyperfine structure but not the direct determinations of μ.

(i) *The Effect of Finite Distribution of Nuclear Charge*

Crawford and Schawlow (1949), following up the original work of Rosenthal and Breit (1932) which had not previously been tested, have considered the effect of a point magnetic dipole situated at the centre of a finite size nucleus for the cases (*a*) where the nuclear charge is distributed uniformly over the surface of a sphere, and (*b*) where the charge is distributed uniformly throughout the volume of a sphere. The radius of the nuclear sphere was calculated by formula (16). The interaction energy between an optical electron and the nuclear magnetic dipole is proportional to the integral

$$I=\int_0^\infty \phi_1\phi_2 y^{-2}\,dy, \qquad \ldots\ldots(20)$$

where ϕ_1, ϕ_2 are the Darwin–Gordon radial functions for the optical electron (notation of Rosenthal and Breit 1932), and $y=2Zr/a_H$, r being measured from

* If I is unknown, it can be obtained by applying the frequency $1{\cdot}400H/F$ Mc/s. to the beam for various assumed values of I: $\frac{1}{2}, \frac{3}{2}\ldots$ or $0, 1, 2\ldots$ etc. Measurement of the frequency which produces resonance between adjacent Zeeman levels for a given H determines F, in particular F_{max}, and hence I.

the point dipole. $\phi_1\phi_2$ in (20) is most heavily weighted for small values of y, where the relativistic wave functions ϕ_1 and ϕ_2 are most distorted, for a finite nuclear charge distribution, from values appropriate to a point charge and Coulomb field. The value of I^0/I was computed, I^0 being the value of I for a nucleus with the charge concentrated at a point. By applying the appropriate correction factor I^0/I to the μ extracted from the hyperfine interval factors, a_{ns}, by the usual spectroscopic calculations (which assume a nucleus with a point charge as well as a point magnetic dipole), Crawford and Schawlow obtain the results of Table 7.

Table 7. Nuclear Moment of ^{205}Tl

(1)	(2)	(3)	(4)	(5)	(6)
Tl III	$7s\ ^3S_{1/2}$	1·348	1·37	1·60	1·56
	$8s\ ^3S_{1/2}$	0·565	1·38	1·61	1·58
	$9s\ ^3S_{1/2}$	0·295	1·39	1·63	1·59
Tl II	6s5g, 6s	5·88	1·44	1·68	1·64
Tl I	$6p\ ^2P_{1/2}$	0·710	1·49	1·57	1·55
	$7s\ ^3S_{1/2}$	0·400	1·35	1·58	1·54
Average				1·61	1·58
Radio-frequency induction			1·628		

(1) Spectrum; (2) term; (3) interaction constant a_{ns} (cm^{-1}); (4) uncorrected μ; (5) corrected μ (nuclear magnetons), charge on surface; (6) corrected μ, uniform volume charge.

A further correction to the 'corrected μ' of Table 7 will be due on account of a finite distribution of magnetic moment. This cannot be exactly calculated at present for Tl, but Crawford and Schawlow point out that the distribution of magnetic moment could reduce the quantity $\phi_1\phi_2 y^{-2}$ to zero for $y<y_0$. In this case the maximum effect would be produced and the corresponding upper limit to the change in I^0/I for an s-electron would be a 3% increase on the values used for Table 7. Thus, either the uniform charge density or uniform surface charge model of the nucleus brings the hyperfine structure μ into agreement with the value determined by Poss (1947) by direct nuclear induction. This does not apply, however, if the spectroscopic μ is corrected in a manner consistent with a deep potential well model of the nucleus (which, incidentally, also gives much worse agreement between observed and calculated isotope shifts than does the uniform charge density model). The effect of the finite distribution of nuclear charge decreases with decreasing atomic number, and for light elements it becomes of the same order of magnitude as the effect due to the finite distribution of magnetic moment. But, as Crawford and Schawlow point out, the 'Rosenthal–Breit' correction appears to be significant for intermediate elements. Thus for Cs Millman and Kusch (1940), by a very accurate magnetic deflection measurement, obtained $\Delta\nu(6s\ ^2S_{1/2})=0{\cdot}30665$ cm^{-1}. This gives the nuclear gyromagnetic ratio $g_I=0{\cdot}7025$ by the usual reduction with the Goudsmit–Fermi–Segrè formula (Fermi and Segrè 1933, Goudsmit 1933): the direct nuclear magnetic resonance measurement of Kusch, Millman and Rabi (1939) (atomic beam method) and Chambers and Williams (1949) (solution of CsCl) gives $g_I=0{\cdot}7315$, i.e. 3·9% higher. The corrected hyperfine structure value gives

$g_I = 0{\cdot}728$, or only 0·4% different from the direct measurement. Similarly, from optical hyperfine structure measurements on La, Crawford and Grace (1935) find $\Delta\nu(6s\,^2S_{1/2}\,La\,III) = 1{\cdot}076\,cm^{-1}$, giving $g_I = 2{\cdot}65$ against 2·761 (4·2% higher) by the direct induction measurement (Chambers and Williams 1949). The Rosenthal–Breit correction to the hyperfine structure causes an increase of 4·1%. For ^{207}Pb, spectroscopic measurements, according to Schawlow, Hume and Crawford (1949) give $\mu = 0{\cdot}49_7$. The direct measurement of Proctor (1949) gives $\mu = 0{\cdot}588 \pm 0{\cdot}001$ nuclear magnetons, i.e. 18% higher. The Rosenthal–Breit correction increases the spectroscopic value by approximately 15%, and a further increase of approximately 3% is probably due on account of the distribution of the magnetic moment throughout the nuclear volume. Thus, for Pb, Tl, La and Cs, application of the Rosenthal–Breit correction brings the spectroscopic μ into agreement with directly measured value.

(ii) *The Effect of Finite Distribution of Nuclear Magnetic Moment*

More definite evidence than that given above for the existence of an observable effect due to the distribution of magnetic moment over the nucleus has been found by Bitter (1949 b) (see also Millman and Kusch 1940) from a different kind of comparison of direct and hyperfine structure determination of μ. By a direct measurement Bitter (1949 a) finds that for Rb, $\mu(^{87}Rb)/\mu(^{85}Rb) = 2{\cdot}0327 \pm 0{\cdot}001$, whilst the corresponding ratio from hyperfine structure calculated by Millman and Kusch (1940) from the results of refined magnetic deflection measurements (Kusch and Millman 1939) is $2{\cdot}0261 \pm 0{\cdot}0003$. As Bitter (1949 b) remarks, this discrepancy is very probably not due to the Rosenthal–Breit effect. Although, for Rb, the Rosenthal–Breit effect gives rise to a correction to μ of about $2Zr_0/a_H$ ($\simeq 1\%$ for Rb) and the observed discrepancy is 0·3%, a very large difference in the radii of ^{87}Rb and ^{85}Rb would clearly be needed to produce the discrepancy, very much larger in fact than is obtained on the assumption that the nuclear volume is proportional to the mass. Moreover, no isotope shift has been observed for Rb, which also indicates the smallness of the nuclear radius change. The magnetic structure of the nucleus can also give rise to an effect of about $2Zr_0/a_H$. It is very much more probable that the discrepancy originates here because, on the basis of the Schmidt nuclear model, there is no objection to assuming that the wave function for the odd proton is quite different for the two isotopes.

There are also discrepancies in similar comparisons for isotope pairs in other elements: ^6Li, ^7Li (Kusch and Mann 1949), ^{203}Tl, ^{205}Tl (Berman, Kusch and Mann 1950) and ^{69}Ga, ^{71}Ga (Kusch 1950), and hydrogen (see §5(iii)).

There are other cases where experiments of the above type (involving the comparison of magnetic moments of pairs of isotopes by direct and hyperfine structure methods) could, and doubtless soon will, be performed. Bohr and Weisskopf (1950) have considered the interpretation of the results from general theoretical considerations and in particular for Rb, where a large effect occurs. The nuclear moment is regarded as distributed and made up of a spin part and an orbital part. The interaction of each of these with the optical electron is considered separately. A quantity, $\epsilon = (\mu_D - \mu_H)/\mu_D$ (where D and H refer to direct and hyperfine structure measurements of μ) is evaluated. This represents, approximately, the fractional change in the total hyperfine structure due to a finite

distribution of nuclear spin and nuclear orbital magnetism. For comparison with observational material the expression

$$\epsilon(1)-\epsilon(2)\simeq 0{\cdot}3b\,\frac{g_S g_L}{g_S-g_L}\left(\frac{1}{g_I(1)}-\frac{1}{g_I(2)}\right), \qquad \ldots\ldots(21)$$

is given. Here the bracketed (1) and (2) refer to the two isotopes in question, g_I being the total nuclear gyromagnetic ratio and g_S and g_L are the corresponding nuclear spin and nuclear orbital ratios: the value of b, depending only on Z and r_0 (equation (16)), is tabulated for $s_{1/2}$ and $p_{1/2}$ electrons for values of Z between 10 and 90 (Bohr and Weisskopf 1950). The effect of finite distribution of nuclear charge (the uniform volume charge sphere model is used) is included in (21).

Most nuclei possessing spin contain an odd number of nucleons, and in this case it is usually assumed that the g factor for spin is that of the odd particle, i.e. $g_S = g$ (proton) for odd Z and $g_S = g$ (neutron) for even Z. The choice of g_L is less certain. In the Margenau and Wigner nuclear model (1940) the nuclear matter as a whole is involved in the orbital momentum, and this leads to $g_L = Z/A$ in units of $e/2Mc$, where M is the nucleon mass. On the other hand, Schmidt (1937), as remarked earlier, §4(viii), assigns the orbital momentum to the motion of the odd particle, which leads to $g_L = 1$ for odd Z and $g_L = 0$ for even Z. These two models lead to different values of $\epsilon(1)-\epsilon(2)$. For Rb ($Z=37$, $A=85$ and 87) the experimental results give $\epsilon(1)-\epsilon(2) = 0{\cdot}33 \pm 0{\cdot}05\%$ for the ground state ($5s_{1/2}$ electron). For $g_L = Z/A = 0{\cdot}43$, $\epsilon(1)-\epsilon(2) = 0{\cdot}11\%$ by (21) and for $g_L = 1$, $\epsilon(1)-\epsilon(2) = 0{\cdot}29\%$. Thus these considerations, as well as the results of isotope shift investigations, also favour the Schmidt model. It is of interest that (21) gives zero $\epsilon(1)-\epsilon(2)$ for A even and Z odd for the Schmidt model, whereas with $g_L = Z/A$, $\epsilon(1)-\epsilon(2)$ should be more than 1% in some cases, for example in Yb and Hg, as Bohr and Weisskopf remark.

(iii) *The Effect of the Proton Motion in Hydrogen and Deuterium*

For the lightest elements more detailed calculations have been made of the effect on hyperfine structure of the dynamics of individual nuclear particles. For ordinary hydrogen, Breit, Brown and Arfken (1949), following up earlier work, have calculated the small effect arising from the motion of the proton. Brown and Arfken (1949) have estimated the effect of a finite proton radius on the nuclear motion correction. The conclusion is that the hyperfine structure should change in the ratio $I/I_0 = 1 - 3m/M +$ terms of order $\alpha^2 m/M$, where I is the hyperfine structure for a proton mass M, and I_0 is the corresponding value when the proton motion (and the magnetic field produced thereby) is neglected.

To account for the ratios of the hyperfine structure observed in hydrogen and deuterium (Nafe, Nelson and Rabi 1947, Nagle, Julian and Zacharias 1947), Bohr (1948) has shown that it is necessary to consider the compound structure of the deuterium nucleus. In the interior part of the atom the electron moves very quickly compared with the proton and neutron. Thus, instead of being characteristic of a point dipole on a point deuterium nucleus, the observed hyperfine structure should be more like that due to a central magnetic dipole on the proton, with the neutron magnetic moment distributed with a density given by the wave function describing the motion of the neutron relative to the proton. This viewpoint should not be much affected by exchange effects between the proton and neutron, the frequency of which should also be small compared with the local

electron frequency. The results greatly improve the agreement of theory with experiment. In a later theoretical investigation of the D hyperfine structure, Low (1950) concludes that theoretical and experimental uncertainties are at present too large to allow conclusions to be drawn concerning possible deviation of the nucleon magnetic fields from those of simple dipoles.

CONCLUSION

The subjects discussed are now being vigorously pursued, and in practically every case the appearance of improved experimental data and theoretical treatment is imminent.

ACKNOWLEDGMENTS

The writer wishes to thank Mr. W. R. S. Garton, Dr. W. M. Gray, Mr. H. Lee and Sir Owen Richardson for loans or gifts of various reprints, etc., which have been helpful in preparing the report.

REFERENCES

ANDREW, A., and CARTER, W. W., 1948, *Phys. Rev.*, **74,** 838.
ARROE, O. H., 1950, *Phys. Rev.*, **79,** 836.
ARROE, O. H., and MACK, J. E., 1950, *J. Opt. Soc. Amer.*, **40,** 386.
BARTLETT, J. H., 1931, *Nature, Lond.*, **128,** 408.
BARTLETT, J. H., and GIBBONS, J. J., 1933, *Phys. Rev.*, **44,** 538.
BERMAN, A., KUSCH, P., and MANN, A. K., 1950, *Phys. Rev.*, **77,** 140.
BETHE, H., 1947, *Elementary Nuclear Theory* (New York : J. Wiley and Sons, Inc.).
BETHE, H., and BACHER, R. F., 1936, *Rev. Mod. Phys.*, **8,** 82.
BITTER, F., 1949 a, *Phys. Rev.*, **75,** 1236; 1949 b, *Ibid.*, **76,** 150.
BOHR, A., 1948, *Phys. Rev.*, **73,** 1107.
BOHR, A., and WEISSKOPF, V. F., 1950, *Phys. Rev.*, **77,** 94.
BRADLEY, L. C., and KUHN, H., 1948, *Nature, Lond.*, **162,** 412.
BREIT, G., 1932, *Phys. Rev.*, **42,** 348; 1950 a, *Ibid.*, **78,** 470; 1950 b, *Ibid.*, **79,** 891.
BREIT, G., ARFKEN, G. B., and CLENDENIN, W. W., 1950, *Phys. Rev.*, **78,** 390.
BREIT, G., BROWN, G. E., and ARFKEN, G. B., 1949, *Phys. Rev.*, **76,** 1299.
BREIT, G., and WILLS, L. A., 1933, *Phys. Rev.*, **44,** 470.
BRIX, P., and FRANK, H., 1950, *Z. Phys.*, **127,** 289.
BRIX, P., and KOPFERMANN, H., 1947, *Nachr. Ges. Wiss. Göttingen*, **2,** 31; 1949, *Z. Phys.*, **126,** 344.
BRIX, P., and STEUDEL, A., 1950, *Z. Phys.*, **128,** 260.
BROCH, E. K., 1945, *Arch. Math. Naturvidensk.*, **48,** 25.
BROWN, G. E., and ARFKEN, G. B., 1949, *Phys. Rev.*, **76,** 1305.
CHAMBERS, W. H., and WILLIAMS, D., 1949, *Phys. Rev.*, **76,** 461.
CONDON, E. U., 1936, *Phys. Rev.*, **49,** 459.
CONDON, E. U., and SHORTLEY, G. H., 1935, *Theory of Atomic Spectra* (Cambridge : University Press).
CONVEY, H., 1940, *Ph.D. Thesis*, Toronto.
CRAWFORD, M. F., 1935, *Phys. Rev.*, **47,** 768; 1936, *Rep. Prog. Phys.*, **3** (London : Physical Society), p. 369; 1938, *Ibid.*, **5,** p. 227.
CRAWFORD, M. F., and GRACE, N. S., 1935, *Phys. Rev.*, **47,** 536.
CRAWFORD, M. F., GRAY, W. M., KELLY, F. M., and SCHAWLOW, A. L., 1950, *Canad. J. Res.* A, **28,** 138.
CRAWFORD, M. F., and SCHAWLOW, A. L., 1949, *Phys. Rev.*, **76,** 1310.
DAVIS, L., NAGLE, D. E., and ZACHARIAS, J. R., 1949, *Phys. Rev.*, **76,** 1068.
DEMPSTER, A. J., 1948, *Phys. Rev.*, **73,** 1125.
DICKINSON, B. H., 1934, *Phys. Rev.*, **46,** 598.
ELSASSER, W., 1934, *J. Phys. Radium*, **5,** 625.
FEATHER, N., 1948, *Nature, Lond.*, **162,** 412.
FEENBERG, E., 1950, *Phys. Rev.*, **77,** 771.

FEENBERG, E., and HAMMACK, K. C., 1949, *Phys. Rev.*, **75,** 1877.
FERMI, E., and MARSHALL, L., 1947, *Phys. Rev.*, **72,** 1139.
FERMI, E., and SEGRÈ, E., 1933, *Z. Phys.*, **82,** 729.
FIERZ, M., 1947, *Nachr. Ges. Wiss. Göttingen*, **3,** 1.
FOSTER, E. W., 1950, *Proc. Roy. Soc.* A, **200,** 429.
FRED, M., TOMKINS, F. S., and BRODY, J. K., 1949, *Phys. Rev.*, **75,** 1772; 1950, *Ibid.*, **79,** 212.
GEIGER, F. E., 1950, *Phys. Rev.*, **79,** 212.
GOUDSMIT, S., 1933, *Phys. Rev.*, **43,** 636.
GOUDSMIT, S., and BACHER, R. F., 1933, *Phys. Rev.*, **43,** 894.
HARTREE, D. R., 1946–7, *Rep. Prog. Phys.*, **11** (London : Physical Society), p. 113.
HARTREE, D. R., and HARTREE, W., 1934, *Proc. Roy. Soc.* A, **149,** 210.
HUGHES, D. S., 1931., *Phys. Rev.*, **38,** 857.
HUGHES, D. S., and ECKART, K., 1930, *Phys. Rev.*, **36,** 694.
HUME, J. N. P., and CRAWFORD, M. F., 1950, *Phys. Rev.*, **78,** 343 (A).
IVANENKO, D., and RODICHEV, V., 1950, *Dokl. Akad. Nauk., S.S.S.R.*, **70,** 801.
JACKSON, D. A., and KUHN, H., 1939, *Proc. Roy. Soc.* A, **173,** 278.
JAECKEL, B., 1936, *Z. Phys.*, **100,** 513.
JAECKEL, B., and KOPFERMANN, H., 1936, *Z. Phys.*, **99,** 492.
JONES, E. G., 1933, *Proc. Phys. Soc.*, **45,** 501.
KLINKENBERG, P. F. A., 1945, *Physica*, **11,** 327.
KOCH, J., and RASMUSSEN, E., 1949, *Phys. Rev.*, **76,** 1417; 1950, *Ibid.*, **77,** 722.
KÖHLER, P., 1939, *Z. Phys.*, **113,** 306.
KOPFERMANN, H., 1940, *Kernmomente* (Leipzig : Akademische Verlagsgesellschaft M.B.H.).
KOPFERMANN, H., and MEYER, D., 1948, *Z. Phys.*, **124,** 685.
KOPFERMANN, H., and WESSEL, G., 1948, *Nachr. Ges. Wiss. Göttingen*, **6,** 53.
KUSCH, P., 1950, *Phys. Rev.*, **78,** 615.
KUSCH, P., and MANN, A. K., 1949, *Phys. Rev.*, **76,** 707.
KUSCH, P., and MILLMAN, S., 1939, *Phys. Rev.*, **56,** 527.
KUSCH, P., MILLMAN, S., and RABI, I. I., 1939, *Phys. Rev.*, **55,** 1176; 1940, *Ibid.*, **57,** 765.
LANDÉ, A., 1933, *Phys. Rev.*, **44,** 1028.
LOW, F., 1950, *Phys. Rev.*, **77,** 361.
MACK, J. E., 1950, *Rev. Mod. Phys.*, **22,** 64.
MCLAY, A. B., and CRAWFORD, M. F., 1933, *Phys. Rev.*, **44,** 986.
MANNING, T. E., ANDERSON, C. E., and WATSON, W. W., 1950, *Phys. Rev.*, **78,** 417.
MARGENAU, H., and WIGNER, E., 1940, *Phys. Rev.*, **58,** 103.
MAYER, M. G., 1948, *Phys. Rev.*, **74,** 235.
MEGGERS, W. F., 1946, *J. Opt. Soc. Amer.*, **36,** 431 : " Spectroscopy, Past, Present and Future " : Bibliography and review of state of analysis of atomic spectra. Hyperfine structure references included.
MILLMAN, S., and KUSCH, P., 1940, *Phys. Rev.*, **58,** 438.
MOORE, C. E., 1949, *Atomic Energy Levels—as derived from the Analyses of Optical Spectra*: Vol. I, ^{1}H–^{23}V (U.S. Department of Commerce, National Bureau of Standards).
MROZOWSKI, S., 1938, *Z. Phys.*, **108,** 204; 1942, *Phys. Rev.*, **61,** 605.
MUNDIE, L. G., and MEISSNER, K. W., 1944, *Phys. Rev.*, **65,** 265.
MURAKAWA, K., 1950, *Phys. Rev.*, **78,** 480.
NAFE, J. E., NELSON, E. B., and RABI, I. I., 1947, *Phys. Rev.*, **71,** 914.
NAGAOKA, H., and MISHIMA, T., 1930, *Sci. Pap. Inst. Phys. Chem. Res. Tokyo*, **13,** 293; 1934, *Ibid.*, **25,** 223.
NAGLE, D. E., JULIAN, R. S., and ZACHARIAS, J. R., 1947, *Phys. Rev.*, **72,** 971.
NORDHEIM, L., 1949, *Phys. Rev.*, **75,** 1894.
OPECHOWSKI, W., and DEVRIES, D. A., 1939, *Physica*, **6,** 913.
POSS, H. L., 1947, *Phys. Rev.*, **72,** 637.
POUND, R. B., 1948, *Phys. Rev.*, **73,** 112.
PROCTOR, W. G., 1949 a, *Phys. Rev.*, **75,** 522; 1949 b, *Ibid.*, **76,** 684.
RABI, I. I., MILLMAN, S., KUSCH, P., and ZACHARIAS, J. R., 1939, *Phys. Rev.*, **55,** 526.
RACAH, G., 1932, *Nature, Lond.*, **129,** 723.
RAINWATER, L. J., RABI, I. I., and HAVENS, W. W., 1949, *Phys. Rev.*, **75,** 1295.

RASMUSSEN, E., 1940 a, *K. Danske Vidensk. Selsk. Mat.-fys. Medd.*, **18,** No. 5; 1940 b, *Phys. Rev.*, **57,** 840.
RITSCHL, R., and SCHOBER, H., 1937, *Z. Phys.*, **38,** 6.
ROLLIN, B. V., 1948–9, *Rep. Prog. Phys.*, **12** (London : Physical Society), p. 22.
ROSENTHAL, J. E., and BREIT, G., 1932, *Phys. Rev.*, **41,** 459.
SCHAWLOW, A. L., HUME, J. N. P., and CRAWFORD, M. F., 1949, *Phys. Rev.*, **76,** 1876.
SCHMIDT, T., 1936, *Z. Phys.*, **101,** 486; 1937, *Ibid.*, **106,** 358; 1938, *Ibid.*, **11,** 332; 1940, *Naturwissenschaften*, **28,** 565.
SCHÜLER, H., and JONES, E. G., 1932 a, *Z. Phys.*, **74,** 631; 1932 b, *Ibid.*, **75,** 563; 1932 c, *Ibid.*, **76,** 14.
SCHÜLER, H., and KEYSTON, J. E., 1931, *Z. Phys.*, **72,** 423.
SCHÜLER, H., and KORSCHING, H., 1937, *Z. Phys.*, **105,** 168; 1938, *Ibid.*, **111,** 386.
SCHÜLER, H., and SCHMIDT, T., 1934, *Z. Phys.*, **92,** 148; 1935, *Ibid.*, **94,** 457.
SCHÜLER, H., and WESTMEYER, H., 1933, *Z. Phys.*, **81,** 565; *Ibid.*, **82,** 685.
SHENSTONE, A. G., and RUSSELL, H. N., 1932, *Phys. Rev.*, **39,** 415.
STERNHEIMER, R., 1950, *Phys. Rev.*, **80,** 102.
TOLANSKY, S., and LEE, E., 1937, *Proc. Roy. Soc.* A, **158,** 110.
TOWNES, C. H., FOLEY, H. M., and LOW, W., 1949, *Phys. Rev.*, **76,** 1415.
VINTI, J. P., 1939, *Phys. Rev.*, **56,** 1120.
WEFELMEIER, W., 1939, *Ann. Phys., Lpz.*, **36,** 373.

A CRITICAL SURVEY OF IONOSPHERIC TEMPERATURES

By N. C. GERSON

Air Force Cambridge Research Laboratories, Massachusetts

CONTENTS

§ 1. Introduction 318
§ 2. Basic ionospheric equations 320
§ 3. Model atmosphere 322
§ 4. Temperature and collisional frequency 324
4.1. Experimental data 324
4.2. Theoretical considerations 325
4.3. Collisional cross section of atomic oxygen 328
4.4. Effective collisional frequency 329
§ 5. Hypsometry and collisional frequency 330
5.1. Theoretical considerations 330
5.2. Effect of neutral particles and ions 331
5.3. Comparison with other investigators 333
§ 6. Temperature and scale height 334
6.1. Scale heights from auroral data 334
6.2. Scale heights determined by other methods 335
§ 7. Temperature and electron concentration 336
7.1. General 336
7.2. Simplifications to Chapman layer 336
7.3. Assumption of a Chapman layer 340
7.4. Conclusions 342
§ 8. Temperature and the recombination coefficient 343
8.1. Recombination at low pressure 343
8.2. Diurnal cycle of recombination coefficient 346
8.3. Limitations of the method 346
8.4. Seasonal F2 region temperature variation 348
8.5. Latitudinal F2 region temperature variations 350
8.6. Conclusion 350
§ 9. Temperature and the influx function 351
9.1. Atmospheric contractive effects 351
9.2. Diurnal chronothermal function 351
9.3. Development of equations 352
9.4. Computational results 353
§ 10. Thermal equilibrium in the mesosphere 357
§ 11. Electron temperatures 358
§ 12. Assumed ionospheric temperatures 359
§ 13. Heating by electrical currents 360
§ 14. Direction of necessary further research 361
Acknowledgments 362
References 362

SYMBOLS

d = radius of active attraction (cm.).
e = charge on electron.
f = frequency of electromagnetic radiation (c/s.).
f_c = critical frequency of probing electromagnetic wave (c/s.).
g = acceleration of gravity (cm/sec^2).
h = Planck's constant (erg./sec.).
k = Boltzmann's constant (erg/° K.).
m = molecular or ionic mass (gm.).
m_e = mass of electron (gm.).
n = number of ions (cm^{-3}).

n_e=number of electrons (cm^{-3}).
q=(Chapman's ionization function) number of electrons produced ($cm^{-3}sec^{-1}$).
t=time (sec.).
u, v=velocity of ions or molecules (cm/sec.).
z=altitude above earth's surface (cm.).
z'=virtual height (cm.).
$H=kT/gMU$=scale height (cm.).
J=intensity of radiation.
M=mean atmospheric molecular weight.
N=number of neutral particles (cm^{-3}).
P=total barometric pressure (dynes/cm^2).
R=radius of earth (cm.).
$\boldsymbol{R}$=gas constant for dry air ($cm^2/°K.\ sec^2$).
T=gas temperature (° K.).
T_e=electron temperature (° K.).
U=atomic mass unit (gm.).
W=ionization potential (ev.).
$Z=(z-z_{max})/H$=altitude in scale height units measured from z_{max}.

α=recombination coefficient (cm^3/sec.).
β=absorption coefficient.
e=base of natural logarithms.
δ=sun's angle of declination.
θ=co-latitude.
η=Stefan's constant.
σ=molecular diameter for collision (cm.).
$\bar{\lambda}$=mean free path (cm.).
χ=sun's zenith angle.
μ=refractive index.
ρ=atmospheric density (gm/cm^3).
ν=collisional frequency between neutral particles (sec^{-1}).
ν'=collisional frequency between electrons and neutral particles (sec^{-1}).
ν''=collisional frequency between electrons and heavy ions (sec^{-1}).
ν_e=collisional frequency between electrons and electrons (sec^{-1}).

CONSTANTS

$e=4{\cdot}8025\times10^{-10}$ E.S. coulomb.
$g_0=980{\cdot}665$ cm/sec^2.
$h=6{\cdot}63\times10^{-27}$ erg./sec.
$k=1{\cdot}3805\times10^{-16}$ erg/mol °K.
$m_e=9{\cdot}107\times10^{-28}$ gm.
$M_0=28{\cdot}979$.
$R=6{\cdot}3676\times10^8$ cm.
$U=1{\cdot}6600\times10^{-24}$ gm.
$e=2{\cdot}7183$.
$\eta=5{\cdot}735\times10^{-5}$ erg/cm^2/sec.
$\rho_0=1{\cdot}2929$ gm/cm^3.

The subscript zero indicates conditions at mean sea level, 45° N.

ABSTRACT. The relative order of importance of the various methods available for the evaluation of temperature from an ionospheric parameter is (*a*) temperature and collisional frequency, (*b*) temperature and auroral scale heights, (*c*) temperature and electron concentration, (*d*) temperature and the influx function, and (*e*) temperature and the recombination coefficient. Method (*e*), still in its infancy, does not yet permit accurate temperature evaluations in the ionosphere. Some indication of a diurnal temperature variation may be obtained from method (*d*).

The probable temperature–altitude relationship in the atmospheric regions 100–400 km., as determined by several techniques, is tabulated below :

Upper Atmospheric Temperatures (° K.)

Altitude (km.)	Method of Determination		
	Collisional frequencies	Auroral scale heights	Electron concentration
100	300	219	—
150	825	531	—
200	1350	1580	—
250	2175	2073	—
300	2400	2455	—
330	—	—	1530–2680
350	3225	2704	—
400	3450	—	—
500	—	—	2200–3910

These values are considered to represent mean annual temperatures at middle latitudes.

The electron and gas temperatures are essentially identical. If the free electrons have an excess of energy (as may result during an ionization process) this energy is quickly distributed among the gas and other particles.

Thermal equilibrium and therefore a Maxwellian velocity distribution of the atmospheric particles exists among the gaseous constituents in the range 100–400 km. The kinetic temperature is therefore equivalent to the gas temperature. Fluctuation phenomena and deviations of the particles from a Maxwellian velocity distribution probably do not occur in the terrestrial atmosphere until much higher altitudes.

A totally ionized atmosphere below 400 km. does not agree with the experimental evidence inasmuch as the temperature deduced on such a basis will not support the observed number density.

Tentatively it has been shown that no large diurnal temperature variations are found in the altitude range 100–400 km. Preliminary estimates of a probable rate of heating based upon the dissipation of excess energy absorbed during an ionization process appear to confirm the preceding statement. However, the rate of heating attributable to non-detectable absorption processes, to large electrical currents in the ionosphere or to the bombardment by large numbers of high energy electrons is unknown.

The existence of temperatures much higher than those given in the above Table would be difficult to explain in the terrestrial atmosphere. Indeed, it may be qualitatively concluded that the maximum atmospheric temperatures occur in the ionospheric regions at an altitude near or somewhat above 400 km. Above this temperature maximum, a general decrease probably occurs to meet the atmospheric boundary temperature with interstellar space. A temperature maximum near the ionospheric regions probably arises from the various absorption or heating processes together with a sufficient number density of particles.

---

§ 1. INTRODUCTION

THOSE physicists who have had occasion to undertake investigations of the upper atmosphere have been impressed not only by the very large number of problems still unsolved but also by the accelerated growth of activity in this field of endeavour. Of the many aspects available for study the temperature in particular has been the subject of a large number of investigations, especially since upon it depend such factors as upper atmospheric pressure, density, molecular mean free path, etc.

Various investigations on upper atmospheric temperature have been reported in the literature, but subsequent summaries have only collected together the various results. The lack of a critical examination of the basic assumptions used in each paper—sometimes entirely different if not contradictory—has been a serious handicap in obtaining a unified temperature–altitude relationship.

In order to present a balanced picture of the present state of knowledge on upper atmospheric temperatures, some references will be made to early work. Initial studies, now largely superseded, were undertaken by Wegener (1911), Chapman and Milne (1920) and Lindemann and Dobson (1923). Later investigations were performed by Maris (1928, 1929), Gowan (1928, 1930) and Angenheister (1932), but again the results were subsequently revised. Following these pioneer studies many investigations were made by diverse methods, the results of which were generally in agreement below an altitude of 100 km. Above this altitude results vary widely from one worker to another, depending upon both the assumptions involved and the specific techniques employed.

Although direct temperature determinations have been achieved with varying degrees of accuracy and success up to only about 100 km., theoretical studies have been made for altitudes up to about 400 km. Some of the methods

used to derive temperatures below 100 km. have been: (*a*) direct soundings with kites, balloons, aircraft and rockets; (*b*) direct observations on the propagation of atmospheric compressional waves; (*c*) studies of the absorption of solar radiation by the ozonosphere; (*d*) mathematical analyses of atmospheric tides and oscillations; (*e*) considerations on the brightness of the twilight sky; (*f*) photographic observations of meteors; (*g*) spectrographic examinations of the light of the night sky and the polar aurora; and (*h*) theories on radiative equilibrium in the upper atmosphere. Results of these observations are well summarized in various papers (Gerson 1949c, Schilling 1950).

For the evaluation of temperatures above 100 km. use has been made almost entirely of: (*i*) deductions from ionospheric parameters, although some values have also been obtained from techniques (*g*) and (*h*). It would be expected that ionospheric factors, determined by radio wave probings of the ionosphere, should provide clues regarding diurnal, annual and geographical variations of temperature, pressure, density, etc., within the layers involved, and the interest in this problem is shown by various monographs concerned with relationships between the ionosphere and upper atmospheric meteorology. Still other treatises (cf. Grimminger 1948, Penndorf 1946) give the summaries of temperature values deduced by various workers.

The various methods available for relating temperature to one ionospheric variable are

(*a*) temperature and collisional frequency,
(*b*) hypsometry and collisional frequency,
(*c*) temperature and scale height,
(*d*) temperature and electron concentration,
(*e*) temperature and the recombination coefficient,
(*f*) temperature and the influx function. Two additional methods, which will not be discussed here also exist, viz. (*g*) temperature equilibrium between incoming radiation, neutral particles and ionization products, (*h*) temperature equilibrium between incoming radiation, neutral particles and dissociation products.

Method (*a*) requires an additional knowledge of either the pressure or the number of particles at the level considered. As neither of these variables is usually known from other sources, the application of (*a*) may be sharply limited. This objection is partially removed when technique (*b*) is used. In this case the collisional frequency, somewhat crudely measured by radio means, rather than the barometric pressure, is used. Since, in most cases, the calculation of scale heights involves the use of electron concentration, process (*c*) must be included under (*d*) unless an independent determination of the scale height is possible. Such a separate evaluation may be made by employing auroral measurements and treating the attenuation of luminosity in a manner very similar to that used in considering the absorption of radiation in the ionosphere. When evaluating temperature from procedure (*d*) several simplifying assumptions are usually made which undoubtedly reduce the final accuracy.

In technique (*e*) the laboratory-observed relationships between the recombination coefficient and the temperatures are applied to the ionosphere. In method (*f*) a diurnal temperature variation is sought which, by producing a nocturnal atmospheric contraction, explains the observed nocturnal ionization densities.

This Report attempts to review critically the various interdependencies which have been proposed between ionospheric parameters and temperature and to obtain better determinations of the temperature in the altitude range 100–400 km. Because an analysis of the upper atmosphere is dependent upon properties of the lower atmosphere, a model atmosphere based upon best available data is chosen as a basis for subsequent investigations (§ 3), and, for completeness, the fundamental equations used later in the text are developed in § 2.

In § 14 are listed specific investigations which must be undertaken in order to resolve outstanding discrepancies in present theories and advance knowledge on upper atmospheric temperatures.

§ 2. BASIC IONOSPHERIC EQUATIONS

Chapman (1931, 1939) has derived expressions showing the functional dependence of temperature upon ionospheric parameters and has calculated the rate of change of ionization density throughout the day. It was assumed that solar ultra-violet ionizing radiation, monochromatic or closely so, falls upon and is absorbed by the atmosphere. The absorption coefficient is assumed constant and a given number of ions n'' are produced by absorption of a unit quantity of radiation. It was further postulated that ionization products are not removed by winds, diffusion, etc., but recombine with one another in accordance with a quadratic law. Under these conditions, the rate of change of ion density is given by

$$dn/dt = q - \alpha n^2 \qquad \ldots\ldots (2.1)$$

under the reasonable assumption of singly ionized particles. The number of ion pairs produced is obtained as follows:

Considering ionizing radiation incident upon the atmosphere at a solar angle χ, the intensity of radiation in penetrating the atmosphere is given by

$$dJ = \beta \rho J \sec \chi \, dz. \qquad \ldots\ldots (2.2)$$

Using the barometric equation $dP = -\rho \boldsymbol{g} dz$ and the perfect gas law $\rho = mP/\boldsymbol{k}T$ equation (2.2) becomes

$$dJ/J = \beta \rho_1 T_1 \sec \chi \, [\exp\{-\int (m\boldsymbol{g}/\boldsymbol{k}T) dz\}] dz/T. \qquad \ldots\ldots (2.3)$$

As the number of ions/cm^3 sec. is given by $Q = \beta n'' J \rho$, (2.3) gives (when the subscript i refers to conditions at a datum level z_1)

$$Q = \beta n'' J_\infty \rho_1 (T_1/T) e^G \qquad \ldots\ldots (2.4)$$

where $$G = -\int (m\boldsymbol{g}/\boldsymbol{k}T) dz + \rho_1 T_1 \beta \sec \chi \int [\exp - \int m\boldsymbol{g}/\boldsymbol{k}T) dz] dz/T$$

and J_∞ is the intensity of radiation at the top of the atmosphere. If a linear temperature lapse rate $T = T_1(1 + bz)$ is assumed, where z is measured from the datum level z_1, and if $\boldsymbol{g}$ is considered a constant,

$$Q = \beta n'' J_\infty \rho_1 (T/T_1)^{-(1+A)} \exp - [\rho_1 \beta H_1 \sec \chi (T/T_1)^{-A}] \qquad \ldots\ldots (2.5)$$

where $A = 1/H_1 b$ and $H_1 = \boldsymbol{k}T_1/m\boldsymbol{g}$. When $b = 0$ an isothermal atmosphere exists and there is obtained the well-known Chapman equation

$$q = \beta n'' J \rho \exp \{-(z/H + \beta \rho H e^{-z/H} \sec \chi)\}. \qquad \ldots\ldots (2.6)$$

In (2.6) $\rho = \rho_1, T = T_1$ and $H = H_1$ and the subscript ∞ to J has been dropped. The equations are valid for non-grazing incidence, generally for angles χ less

than 80°. In the case of (2.6), the height of maximum ion production $z_{\max}$ for a given zenith angle is found to be

$$z_{\max} = H \ln (\beta\rho H \sec \chi). \qquad \ldots\ldots(2.7)$$

The rate of ion production at this height then becomes

$$q = \beta n'' J \rho \exp - [(z_{\max}/H) + 1]. \qquad \ldots\ldots(2.8)$$

Obviously q attains its maximum when $z_{\max}/H = 0$, i.e. when $\beta\rho H = \cos\chi$. Thus, if $q'_{\max}$ is the rate of maximum production of ions,

$$q'_{\max} = \beta\rho n'' J/e = n'' J (\cos\chi)/eH. \qquad \ldots\ldots(2.9)$$

Note that $q_{\max}$, the maximum value of $q'_{\max}$, occurs when $\chi = 0$, so that $\beta\rho H = 1$ and

$$q_{\max} = n'' J/eH. \qquad \ldots\ldots(2.10)$$

From (2.6), (2.9) and (2.10) $\quad q = q_{\max} e^{f(T)} = q'_{\max} e^{F(T)}, \qquad \ldots\ldots(2.11)$

where $\quad f(T) = 1 - Z - e^{-Z} \sec\chi; \; F(T) = 1 - Z - e^{-Z}. \qquad \ldots\ldots(2.12)$

Under equilibrium conditions $dn/dt = 0$, whence, from (2.1),

$$n = (q/\alpha)^{1/2} \qquad \ldots\ldots(2.13)$$

making it possible to write for the isothermal case if α is constant,

$$n = n_{\max} e^{f(T)/2} = n'_{\max} e^{F(T)/2} \qquad \ldots\ldots(2.14)$$

where $n_{\max}$ denotes the electron density at the level of maximum ionization when $\chi = 0$ and $n'_{\max}$ is the electron density at the same level when $\chi \neq 0$.

Considering a linear temperature increase with height, from (2.5),

$$n = (n'' J \rho_1 \beta/\alpha)^{1/2} (T/T_1)^{-(1+A)/2} [\exp \{\rho_1 \beta H_1 \sec\chi (T/T_1)^{-A}\}/2]. \qquad \ldots\ldots(2.15)$$

When $\chi = 0$ and the first three terms of the exponential expansion only are used (2.12) and (2.14) become

$$\tfrac{1}{2} f(T) = -\tfrac{1}{4} Z^2; \; n = n'_{\max} (1 - \tfrac{1}{4} Z^2). \qquad \ldots\ldots(2.16)$$

From (2.9) and (2.13) the ratio between two electron concentrations n_a and n_b under different conditions of rate of production is

$$n_a/n_b = [(J_a H_b \alpha_b \cos\chi_a)/(J_b H_a \alpha_a \cos\chi_b)]^{1/2}. \qquad \ldots\ldots(2.17)$$

For grazing incidence some of the above equations may not be used. Thus instead of (2.6), the equation for q is

$$q = \beta n'' J \rho \exp [-(z/H + \beta\rho D)] \qquad \ldots\ldots(2.18)$$

$$D = \int_{z=\infty}^{z=C\sec\chi - R} e^{-z/H} \{(R+z)/[(R+z)^2 - C^2]^{1/2}\} dz; \qquad \ldots\ldots(2.19)$$

C is constant for any particular incident beam of solar radiation. Equation (2.18), which must be used whenever $\chi \geqslant 80°$, may be written in terms of $q_{\max}$ and $z'_{\max} = H \ln(\beta\rho H)$ as

$$q = q_{\max} e^{\Psi(T)} \qquad \ldots\ldots(2.20)$$

where $\quad \Psi(T) = 1 - (z - z'_{\max})/H - (D/H) \exp (z'_{\max}/H). \qquad \ldots\ldots(2.21)$

The latter equation may be solved by numerical integration. Use will also be made of the refractive index of an ionized gas which is given by

$$\mu^2 = 1 - ne^2/\pi m f^2. \qquad \ldots\ldots(2.22)$$

At the reflection point of a probing electromagnetic wave, where $\mu = 0$,

$$n = (\pi m/e^2) f^2. \qquad \ldots\ldots (2.23)$$

The number of ions is calculated from experimentally determined values of f.

In this work the Sellmeyer theory is adopted in accordance with the theoretical conclusions of Darwin (1943) and Ginsburg (1944 b) and the experimental interpretations of Martyn and Munro (1939).

§ 3. MODEL ATMOSPHERE

In the course of this investigation it will be necessary to refer to a 'model atmosphere'. After careful consideration of the work of Bates (1949), Penndorf (1949), Gerson (1949 c), Grimminger (1948), Jacchia (1948), Hulburt (1947), Warfield (1947), Gutenberg (1946), Gonnermann and Stranz (1944), Whipple

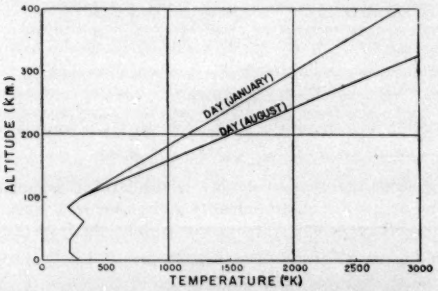


Figure 1. Temperature–altitude relationship in the model atmosphere.

Table 1. Atmospheric Properties up to altitude 400 km.

Altitude (km.)		Temperature (°K.)	Pressure (mb.)	Number density* (cm^{-3})	Density† (gm/cm^3)	Classical mean free path‡ (cm.)
0		288	1013·25	$2{\cdot}549\times10^{19}$	$1{\cdot}226\times10^{-3}$	$8{\cdot}625\times10^{-6}$
10·8		218	$2{\cdot}349\times10^{2}$	$7{\cdot}805\times10^{18}$	$3{\cdot}756\times10^{-4}$	$2{\cdot}816\times10^{-5}$
32		218	8·608	$2{\cdot}860\times10^{17}$	$1{\cdot}376\times10^{-5}$	$7{\cdot}685\times10^{-4}$
62		330	$2{\cdot}041\times10^{-1}$	$4{\cdot}480\times10^{15}$	$2{\cdot}156\times10^{-7}$	$4{\cdot}906\times10^{-3}$
84		200	$1{\cdot}204\times10^{-2}$	$4{\cdot}361\times10^{14}$	$2{\cdot}098\times10^{-8}$	$5{\cdot}041\times10^{-1}$
94		262·5	$2{\cdot}834\times10^{-3}$	$7{\cdot}821\times10^{13}$	$3{\cdot}763\times10^{-9}$	2·811
100		300	$1{\cdot}488\times10^{-3}$	$3{\cdot}593\times10^{13}$	$1{\cdot}429\times10^{-9}$	6·118
300	August	2700	$1{\cdot}142\times10^{-5}$	$3{\cdot}064\times10^{10}$	$1{\cdot}127\times10^{-12}$	$7{\cdot}174\times10^{3}$
	January	2000	$3{\cdot}983\times10^{-6}$	$1{\cdot}443\times10^{10}$	$5{\cdot}307\times10^{-13}$	$1{\cdot}542\times10^{4}$
400	August	3900	$5{\cdot}692\times10^{-6}$	$1{\cdot}057\times10^{10}$	$3{\cdot}619\times10^{-13}$	$2{\cdot}079\times10^{4}$
	January	2850	$1{\cdot}546\times10^{-6}$	$3{\cdot}929\times10^{9}$	$1{\cdot}345\times10^{-13}$	$5{\cdot}594\times10^{4}$

* From $N = P/kT$. † From $\rho = PM/M_0RT$. ‡ From $\bar{\lambda}^{-1} = \sqrt{2}\pi N\sigma^2$; $\sigma = 3{\cdot}2\times10^{-8}$ cm.

(1943), Penndorf and Flohn (1942), Knight (1936) and Diehl (1925) the temperature–altitude functions shown in Figure 1 and Table 1 were adopted for this investigation (see also Roberts 1949, Nazarek 1950, Regula 1949). Temperatures of the model atmosphere lie on the upper limit of values obtained by other investigators, and are considered to hold at latitude 45° N. for January and

August conditions respectively. A suggestion has been made (Weeks and Wilkes 1947) that a temperature decrease occurs above the E region. As, however, the results are far from conclusive, this possibility is not considered here.

By the use of the hydrostatic equation, considering (*a*) a dissociation of molecular oxygen or nitrogen to occur linearly within a given stratum and (*b*) a value of gravity which varies inversely as the square of the distance from the centre of the earth, it is possible to compute corresponding pressures. These values are also given in Table 1. It will be assumed that oxygen dissociates in the region 94–100 km., and is entirely in the atomic state above 100 km. (Penndorf 1949, Rakshit 1947, Wulf and Deming 1938). Although nitrogen dissociation at great altitudes is regarded as an accepted fact, the abundance of the atomic form at any specific level is not known conclusively. The high-frequency solar ultra-violet radiations arriving at these levels may be insufficient for direct dissociation, but by predissociation of N_2, the photodissociation of N_2^+, etc., atomic nitrogen in substantial amounts may be formed (Ta-You-Wu 1944, Herzberg 1948, Bernard 1939).

The concentration of nitrogen–oxygen compounds such as N_2O, NO, etc., is considered to be small. Experimentally, the relative concentration of atomic

Table 2. Mean Atmospheric Molecular Weight at Various Altitudes

Molecular weight	Altitude (km.)	Percentage dissociation	
		O_2	N_2
$M=M_0=28{\cdot}98$	0–94	0	0
$M=M_0[1-c_3(z-c_3)]^{-1}$	94–100	0–100	0
$M=M'=23{\cdot}95$	100–200	100	0
$M=M'[1-c_4(z-c_5)]^{-1}$	200–400	100	0–25
$M=20{\cdot}61$	400	100	25

$c_2=3{\cdot}5\times10^{-7}$/cm.; $c_3=9{\cdot}4\times10^{6}$ cm.; $c_4=8{\cdot}09\times10^{-9}$/cm.; $c_5=2\times10^{7}$ cm.

nitrogen has been interpreted from auroral spectrographs to increase with height (Dufay 1943). Bernard (1941) estimated from photographic plates of the aurorae that the fraction of nitrogen dissociation increases from about 0·07 near 100 km. to about 0·75 near 600–700 km. However, as there may be reason for suspecting that the percentage of atomic nitrogen in the auroral zones is greater than that found at lower latitudes, Bernard's results may not have been applicable over the whole globe.

In any event while the theory and experiment agree in allowing limited amounts of atomic nitrogen, complete dissociation of nitrogen cannot be accepted. For example, interpretation of the auroral band spectra to extreme altitudes requires the presence of considerable amounts of the molecular forms N_2 and N_2^+ (Barbier 1947, Vegard 1946, Vegard and Kvifte 1945).

George (1947) assumed nitrogen dissociation to begin below 160 km., becoming, at 400 km., 80% and 15% complete during daylight and darkness, respectively. However, in view of the preceding information such assumptions seem somewhat arbitrary. In our model atmosphere we postulate no nitrogen dissociation to 200 km., thence a dissociation increasing linearly with altitude to 25% at 400 km.

Although Mitra and Rakshit (1938) found that complete mixing with no gravitational separation of the atmospheric constituents may reasonably exist to 300–350 km., it should be recalled that these workers propounded a temperature

rise of only 4°K/km., other conditions being about as given in the model atmosphere. In the present work, complete mixing of the atmospheric constituents to 400 km. is adopted. Under the above circumstances, the mean molecular weight of the atmosphere has the values given in Table 2.

The atmospheric parameters at 300 km. and 400 km., tabulated in Table 3, include pressure, classical mean free path and number of particles under given temperature conditions. A linear temperature gradient is assumed from $T = 300°$ K. at 100 km. to the given temperatures at the specified altitude.

Table 3. Values of Pressure, Number Density and Mean Free Path at 300 km. and 400 km. under various assumed temperatures (determined by use of the hypsometric equation assuming a linear temperature gradient from $T = 300°$ K. at $z = 100$ km.)

T (° K.)	P (mb.)	N (cm^{-3})	$\bar{\lambda}$ (cm.)
$z = 300$ km.			
300	$4{\cdot}238 \times 10^{-11}$	$1{\cdot}023 \times 10^{8}$	$2{\cdot}148 \times 10^{3}$
500	$2{\cdot}281 \times 10^{-9}$	$3{\cdot}305 \times 10^{7}$	$6{\cdot}652 \times 10^{4}$
800	$4{\cdot}850 \times 10^{-8}$	$4{\cdot}392 \times 10^{8}$	$5{\cdot}005 \times 10^{3}$
1200	$4{\cdot}328 \times 10^{-7}$	$2{\cdot}613 \times 10^{9}$	$8{\cdot}413 \times 10^{4}$
2000	$3{\cdot}983 \times 10^{-6}$	$1{\cdot}443 \times 10^{10}$	$1{\cdot}524 \times 10^{4}$
2700	$1{\cdot}142 \times 10^{-5}$	$3{\cdot}064 \times 10^{10}$	$7{\cdot}174 \times 10^{3}$
3000	$1{\cdot}589 \times 10^{-5}$	$3{\cdot}837 \times 10^{10}$	$5{\cdot}729 \times 10^{3}$
4000	$3{\cdot}557 \times 10^{-5}$	$6{\cdot}442 \times 10^{10}$	$3{\cdot}412 \times 10^{3}$
5000	$6{\cdot}087 \times 10^{-5}$	$8{\cdot}819 \times 10^{10}$	$2{\cdot}493 \times 10^{3}$
10,000	$2{\cdot}166 \times 10^{-4}$	$1{\cdot}569 \times 10^{11}$	$1{\cdot}401 \times 10^{3}$
$z = 400$ km.			
300	$2{\cdot}230 \times 10^{-14}$	$5{\cdot}385 \times 10^{5}$	$4{\cdot}082 \times 10^{11}$
600	$3{\cdot}645 \times 10^{-11}$	$4{\cdot}401 \times 10^{8}$	$4{\cdot}995 \times 10^{8}$
1050	$4{\cdot}105 \times 10^{-9}$	$2{\cdot}832 \times 10^{7}$	$7{\cdot}762 \times 10^{4}$
1650	$8{\cdot}676 \times 10^{-8}$	$3{\cdot}809 \times 10^{8}$	$5{\cdot}771 \times 10^{3}$
2850	$1{\cdot}546 \times 10^{-6}$	$3{\cdot}939 \times 10^{9}$	$5{\cdot}594 \times 10^{4}$
3900	$5{\cdot}692 \times 10^{-6}$	$1{\cdot}057 \times 10^{10}$	$2{\cdot}079 \times 10^{4}$
4350	$8{\cdot}501 \times 10^{-6}$	$1{\cdot}416 \times 10^{10}$	$1{\cdot}553 \times 10^{4}$
5850	$2{\cdot}230 \times 10^{-5}$	$2{\cdot}761 \times 10^{10}$	$7{\cdot}960 \times 10^{3}$
7350	$4{\cdot}194 \times 10^{-5}$	$4{\cdot}133 \times 10^{10}$	$5{\cdot}318 \times 10^{3}$
10,000	$8{\cdot}592 \times 10^{-5}$	$6{\cdot}224 \times 10^{10}$	$3{\cdot}532 \times 10^{3}$

4. TEMPERATURE AND COLLISIONAL FREQUENCY

4.1. *Experimental Data*

Before discussing the theoretical aspects of collisional frequency it will prove advantageous to present the experimental results. These values as given by various workers are shown in Table 4 and Figure 2. Several methods have been used for measuring collisional frequency: for the E region, use has been made of the phenomenon of cross modulation involving the interaction of radio waves. In general, however, measurements of absorption of radio waves in the ionosphere yield an amplitude reflection coefficient as a function of the probing radio frequency. Under certain assumptions these quantities may then be related to the collisional frequency of the particles at the level explored.

4.2. *Theoretical Considerations*

As the upper atmosphere consists of a gaseous mixture of neutral particles, heavy ions and electrons, collisional frequencies must be explicitly specified, whether of neutral–neutral particles, neutral particles–electrons, ions–electrons, or electrons–electrons, etc. Each action is a distinct physical process, described

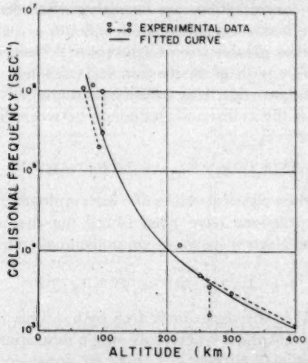


Figure 2. Collisional frequencies as a function of altitude as deduced from ionospheric observations.

Table 4. Experimentally Determined Values of Collisional Frequency

Altitude (km.)	Collisional frequency (sec^{-1})	Authority
70–95	$1{\cdot}1\times10^{6}-2\times10^{5}$	Ratcliffe and Shaw (1948)
85	$1{\cdot}2\times10^{6}$	Huxley, Foster and Newton (1948)
100	$1{\cdot}0\times10^{6}-3\times10^{5}$	Chapman and Bartels (1940)
100	$1{\cdot}9\times10^{5}$	Bailey and Martyn (1935)
220	$1{\cdot}2\times10^{4}$	Toshniwal, Pant and Bajpai (1936)
250–400	$5\times10^{3}-1{\cdot}2\times10^{3}$	Vilbig (1944)
		Farmer and Ratcliffe (1935 b)
265	$3{\cdot}6\times10^{3}-1{\cdot}6\times10^{3}$	Eckersley (1935)
		Farmer and Ratcliffe (1935 a)
300	3×10^{3}	Ginsburg and Alpert (1944)

by a specific relationship. Several of these dependences will be discussed below.

In a gas having a Maxwellian velocity distribution, the average speed and the mean free path are respectively (neglecting external force fields):

$$\bar{u}=(8kT/\pi m)^{1/2} \qquad \ldots\ldots(4.1)$$

$$\bar{\lambda}=(3/\pi)^{1/2}/4\sigma^{2}N. \qquad \ldots\ldots(4.2)$$

Since the mean collisional frequency is defined by $\nu=\bar{u}/\bar{\lambda}$ where ν represents an average over all particle velocities,

$$\nu=8\sigma^{2}N(2kT/3m)^{1/2}. \qquad \ldots\ldots(4.3)$$

Equation (4.3) may be used for the determination of temperatures only if both the total number of atoms and molecules N and their mean molecular mass

are known. If the pressure rather than the total number of particles is chosen as one of the variables (4.3) becomes

$$\nu = 8P\sigma^2(2/3m\boldsymbol{k}T)^{1/2} \qquad \ldots\ldots(4.4)$$

where P is the partial pressure of the atoms and molecules.

Since equations (4.3) and (4.4) are valid for collisions between the neutral particles only, further considerations are necessary when dealing with electrons. Because of the smaller mass, the collisional frequency ν' between electrons and neutral particles becomes greater by a factor $(m/m_e)^{1/2}$ over that given in (4.4). Also, since the mean free path of an electron is $4\sqrt{2}$ longer than the mean free path of uncharged gas particles, the collisional frequency ν' must be divided by this amount. Thus the collisional frequency between electrons and neutral particles is, assuming $T = T_e$

$$\nu' = 2\sigma^2 N(\boldsymbol{k}T/3m_e)^{1/2} \; ; \; \nu' = 2\sigma^2 P/(3m_e\boldsymbol{k}T)^{1/2}. \qquad \ldots\ldots(4.5)$$

In general, $\nu'/\nu \sim 70$ when classical values of σ are employed.

Quite different expressions have been found for the collisional frequency between ions. For the electron-positive ion collisional frequency ν'' Majumdar (1937) derived

$$\nu_M'' = (4\pi e^4/3)/(2\pi m_e\boldsymbol{k}^3)^{1/2}(n/T^{3/2})[\ln(T\bar{\lambda}''^2) - B] \qquad \ldots\ldots(4.6)$$

where $B = 24{\cdot}58$ and $\bar{\lambda}''$ is the mean ionic free path. This relationship is based upon an electron–ion ionosphere where only single ionization of the constituents occurs. For computational purposes, $\bar{\lambda}''$ may be considered equal to $\bar{\lambda}/5$. A similar expression for the collisional frequency between electrons and singly ionized ions was obtained by Ginsburg (1944 a):

$$\nu_G'' = [\pi e^4 n(3/m_e)^{1/2}/(\boldsymbol{k}T)^{3/2}]\ln(\boldsymbol{k}Te^2n^{1/3}). \qquad \ldots\ldots(4.7)$$

It will be found below that (4.6) and (4.7) give results which are of the same order of magnitude. For the electron–neutral particle collisional frequency Ginsburg employed essentially (4.5).

For completeness, the mean collisional frequency ν_e between electrons should also be considered. This quantity as derived by Drukarev (1946) is

$$\nu_e = [n_e e^4/m_e^{1/2}(\boldsymbol{k}T_e)^{3/2}]\ln[(\boldsymbol{k}T_e)^{3/2}/e\bar{\lambda}'^2(8\pi n_e)^{1/2}] \qquad \ldots\ldots(4.8)$$

where $\bar{\lambda}' = 4\sqrt{2}\bar{\lambda}$.

It should be recalled that the collisional frequency as measured experimentally in the ionosphere is the sum of the various collisional processes:

$$\nu_{eff} = \nu' + \nu'' + \nu_e. \qquad \ldots\ldots(4.9)$$

Figures 3 and 4 show the relative magnitudes and the variation with temperature of each component of ν_{eff} at 300 km. and 400 km. respectively. For the calculations the results of Table 3 were used, assuming in (4.6) and (4.7) that $n = N/2$, i.e. that an entirely ionized atmosphere existed. In evaluating (4.5), two general cases were considered: (*a*) σ = constant (for nitrogen), and (*b*) $\sigma = \sigma(T)$ (for atomic oxygen). Several values of $\sigma(T)$ (see § 4.3 for details) were used to permit intercomparison of results. The near equality of ν_M'' and ν_G'' should be noted.

With his evaluation of (4.8) Drukarev (1946) obtained the values $\nu_e = 30$/sec. and $\nu_e = 3$/sec. for the F2 and E layers respectively. Drukarev apparently

considered a mean free path of the order of 10^{-8} cm., which is typical of sea-level atmospheric conditions. From Tables 1 and 3, however, it is seen that the mean free path for the F2 region is about 10^4 cm. and for the E region about 10 cm. When these values are inserted in (4.8), the logarithm of a negative number occurs. Indeed, in the terrestrial atmosphere it is only for conditions near the surface of the earth, where $\bar{\lambda}'$ is very small, that a real solution for ν_e may be obtained from Drukarev's formula.

It is clear from the results that under the given assumptions the collision of electrons with ions is greater than collision of electrons with neutral particles by a factor of about 10^4 or 10^5 at temperatures near 1,000° K. Furthermore, since the observed collisional frequency at 300 km. is of the order of 10^3/sec., a totally

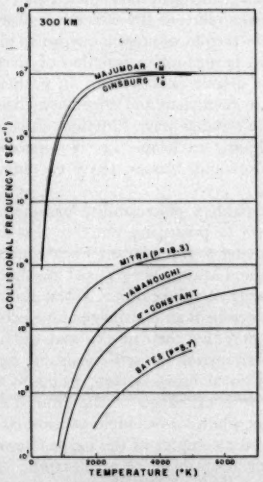


Figure 3.

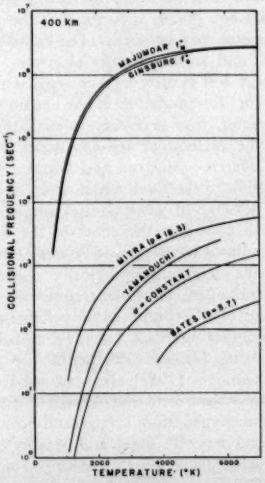


Figure 4.

Computed values of collision frequencies at 300 km. (Figure 3) and 400 km. (Figure 4). The terms 'Majumdar' and 'Ginsburg' refer to electron–ion collisions in a totally ionized atmosphere; 'Mitra', 'Yamanouchi' and 'Bates' to electron–atomic oxygen collisions for various states of the oxygen atom; and 'σ=constant' to electron–nitrogen (both atomic and molecular) collisions.

ionic atmosphere would require a temperature of about 200° K. whereas a free electron–neutral particle atmosphere would imply temperatures of about 4,000° K. if solely nitrogen particles (σ=constant) were considered, but 1,300–7,000° K. considering atomic oxygen (under various hypothetical collisional cross sections). With regard to an ionosphere composed entirely of ions, it is at once obvious that the low temperature of 200° K. at 300 km. is inadmissible inasmuch as it would not allow the existence of the measured number of ionized particles at this level (see Table 3). *An ionosphere below 400 km. having all particles ionized may therefore be rejected immediately.*

In many instances it has been assumed that ion–electron collisions in the ionosphere were many orders of magnitude greater than neutral particle–electron collisions. Such results are not found here. The conclusions are in accord with those obtained by Cowling (1945), who determined that when $N/n > 10^6$, collisions between ions of opposite sign are comparatively unimportant.

Several remarks regarding the use of (4.5) are in order. Since the temperature varies as the fourth power of the collisional diameter*, careful consideration must be given to the value of σ adopted for the computations. Obviously even a small variation from the true value of σ will result in a severalfold change of T.

4.3. *Collisional Cross Section of Atomic Oxygen*

For slow electrons interacting with oxygen or nitrogen molecules, Fisk (1936) and Loeb (1934) indicate that the gas kinetic value of the effective collisional diameter may be used. The collisional cross section of atomic oxygen to about 1 ev. and smaller energy electrons, however, is not only a function of electron energy but is also very dependent upon the atomic field forces. In particular, for the determination of the elastic scattering coefficient and effective collisional diameter, it is necessary to evaluate Hartree's radial wave functions calculated by the method of self-consistent field including exchange, i.e. it is necessary (cf. Hartree, Hartree and Swirles 1939, Mott and Massey 1949) to obtain a solution of the Foch equations.

Because of uncertainties in the atomic field, a polarizability parameter p, which considers the effect of the 3s electron in polarizing the core, has been introduced by Bates and Massey (1943). Using p, the relevant Foch equations may be integrated numerically. Later, Bates and Massey (1947) used this method and obtained collisional frequencies for a range of values of p. Mitra, Ray and Ghosh (1940) computed the collisional cross section of atomic oxygen for electrons by quantal methods, but these authors only note that their results were obtained by using Hartree's technique for the determination of self-consistent fields. Yamanouchi (1947) also evaluated the collisional cross section, using for the mean atomic field the self-consistent field, with exchange, of the ground state of the oxygen atom. It should be noted that while Yamanouchi considered the ground state ^{3}P, Bates and Massey considered a complex of the oxygen atom in the ground state plus a free s electron (yielding an atom in a ^{2}P or ^{4}P state).

Figure 5 is a graph of collisional diameter plotted against temperature for oxygen atoms interacting with slow electrons, deduced from the above theories. Two values were taken from Bates and Massey's (1947) work; one corresponding to a polarizability $p = 5{\cdot}7$, the predicted value, and the other to $p = 18{\cdot}3$, an almost resonance effect. It is evident that the cross section for elastic scattering is highly dependent upon the polarization factor. It is also apparent from Figure 5 that the curve obtained from Mitra and collaborators agrees closely with that deduced by Bates and Massey for $p = 18{\cdot}3$ atomic units.

With such various and divergent values of $\sigma = \sigma(T)$ for the effective collisional diameter of oxygen atoms with slow electrons, confusion exists as to the value of σ to be chosen. A polarization of $p = 18{\cdot}3$ corresponds very closely to resonance for attachment of the electron to the oxygen atom. From a consideration of the physical processes involved in the upper atmosphere, however, it seems improbable

* The collisional or effective mean diameter is defined as $(S/\pi)^{1/2}$ where S=collisional cross section (cm^2).

that such large polarizations of the core are common. Furthermore, as will be shown below, the collisional frequencies calculated when $p = 18{\cdot}3$ are much greater than those observed. On the other hand, the collisional frequencies evaluated when $p = 5{\cdot}7$ are so low that excessive temperatures are demanded in the higher atmospheric region. Yamanouchi's values appear to be the best available.

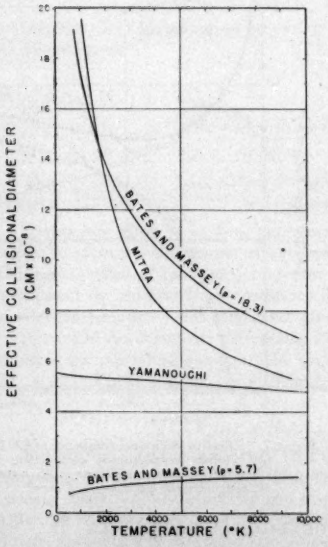


Figure 5. Effective collisional diameter of atomic oxygen with electrons as a function of temperature as deduced by various investigators.

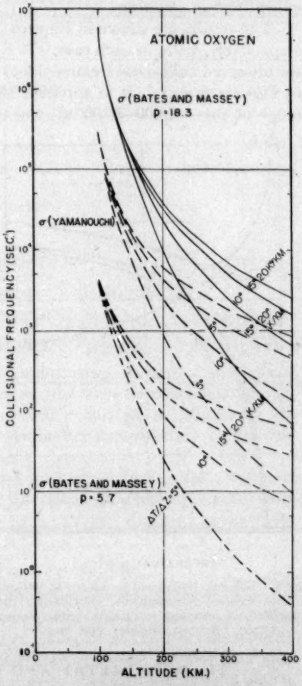


Figure 9. Computed collisional frequencies of electrons with atomic oxygen as a function of altitude considering both various temperature lapse rates and various states of the atom.

4.4. *Effective Collisional Frequency*

Since it has been shown that ν_e may be neglected in the terrestrial atmosphere, and that $\sigma = \sigma(T)$ for atomic oxygen, equation (4.9) should be rewritten as

$$\nu_{\text{eff}} = \nu'' + \nu'_{N_2+N} + \nu_O' \qquad \ldots\ldots (4.10)$$

where

$$\nu'_{N_2+N} = 2\sigma^2 N_{N_2+N} (kT/3m_e)^{1/2} \qquad \ldots\ldots (4.11)$$

(N_{N_2+N} = number of nitrogen molecules and atoms) and

$$\nu_O' = 2[\sigma'(T)]^2 N_O (kT/3m_e)^{1/2} \qquad \ldots\ldots(4.12)$$

(N_O = number of oxygen atoms).

In equation (4.10), ν'' was evaluated with $n = 2{\cdot}5 \times 10^6/\text{cm}^3$ at 300 km., and $n = 1 \times 10^6/\text{cm}^3$ at 400 km., and with the appropriate number density for the constituent in question.

The results are shown in Figures 6 and 7 for altitudes of 300 km. and 400 km. respectively, with, in each case, $\nu'' = \nu_M''$ and Yamanouchi's value for $\sigma(T)$. With the observed collisional frequencies at 300 km. and 400 km. from Figure 2 applied to Figures 6 and 7, it is apparent that temperatures at 300 km. lie within the range of about 2,000–3,100° K., and temperatures at 400 km. lie below 4,100° K.

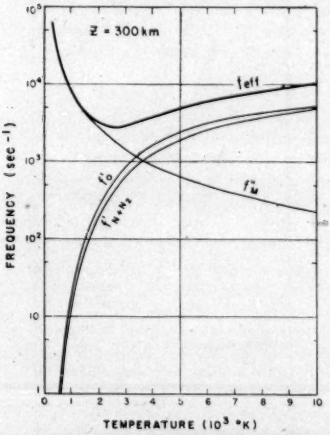


Figure 6. Effective collisional frequency at 300 km. for various temperatures, considering the effect of electron collisions with atomic oxygen, nitrogen particles, and ions.

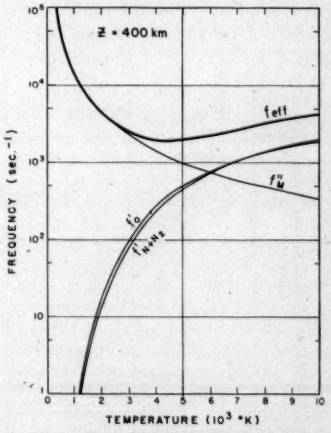


Figure 7. Effective collisional frequency at 400 km. for various temperatures, considering the effect of electron collisions with atomic oxygen, nitrogen particles, and ions.

§5. HYPSOMETRY AND COLLISIONAL FREQUENCY

Note: Section 4, Temperature and Collisional Frequency, and this section treat a single problem but from two distinct aspects. In the first instance, the collisional frequency at a constant geometric surface is examined, considering temperature as a variable. In this section, however, the collisional frequency is studied as a function of altitude considering various assumed temperature lapse rates. Because of the extremely interesting results, it is deemed worth while to present both analyses.

5.1. *Theoretical Considerations*

Vegard (1939) and more recently Woodward (1948) have suggested that the pressure of the upper atmosphere may be materially influenced by coulomb and radiational forces as well as by gravitation. Coulomb forces would arise from the interaction of possible large-scale space charges in the terrestrial atmosphere,

whereas the radiational forces would be those exerted by solar and possibly by stellar radiation fields. The first effect could conceivably arise from a suitable charge separation process which might be effective at great altitudes. However, Spitzer's (1949) calculations have clearly indicated that such a hypothesis is untenable in view of the energies required. Also, while radiational forces may be important at extreme distances from the earth, they may be ignored in the altitude range under consideration. Thus this section is concerned only with the effect of gravitational forces upon atmospheric pressure.

From the hypsometric equation

$$dP/P = -(mg/kT)dz. \qquad \ldots\ldots(5.1)$$

A linear temperature–height function will be postulated of the type

$$T = T_1[1 - a(z - z_1)]. \qquad \ldots\ldots(5.2)$$

Integration of (5.1) accepting m and g as average values in the interval of integration gives

$$P/P_1 = (T_1/T)^{mg/akT_1}. \qquad \ldots\ldots(5.3)$$

Substituting (4.5) in (5.3)

$$\nu'/\nu_1' = (\sigma/\sigma_1)^2(T_1/T)^{mg/akT_1+1/2} \qquad \ldots\ldots(5.4)$$

Eliminating the coefficient of temperature increase with altitude

$$\nu'/\nu_1' = (\sigma/\sigma_1)^2(T_1/T)^x \qquad \ldots\ldots(5.5)$$

where

$$x = mg(z - z_1)/k(T - T_1) + 1/2. \qquad \ldots\ldots(5.6)$$

Evaluations from (5.5) were performed for the cases of neutral atomic and molecular nitrogen, and atomic oxygen, taking for the latter several values of $\sigma(T)$. The initial value of collisional frequency ν_1' was evaluated at 100 km. for the constituent in question from conditions in the model atmosphere. In the case of nitrogen particles σ was considered constant.

As would be expected, the results found here agree in general with those found in the preceding section. The slight discrepancies which appear between the results of both sections are attributable to the slightly different assumptions made.

5.2. *Effect of Neutral Particles and Ions*

Results for nitrogen are given in Figure 8 for several different values of temperature lapse rate $\Delta T/\Delta z$. The datum collisional frequency ν_1' is that existing solely between electrons and nitrogen particles at the 100 km. altitude. Although a value of $\Delta T/\Delta z = 20°$ K/km. seems to fit the observational data most closely, acceptance of this lapse rate would imply temperatures of about 6,300° K. at 400 km. in the terrestrial atmosphere. Obviously since a planetary and not a stellar atmosphere is under consideration, such magnitudes of temperature are completely inadmissible.

To show conclusively the most probable value to be chosen, the several proposed $\sigma(T)$ relationships were used in evaluating (5.5) for atomic oxygen (see Figure 9). It should be observed that collisional frequencies obtained when $p = 5{\cdot}7$ do not fit the experimental evidence. Even with the large rate of temperature increase with height of 20° K/km., collisional frequencies in this instance are about 10^{-2} times those obtained for nitrogen, and, as will be shown below, for ions also. When $p = 18{\cdot}3$ the computed magnitude of ν_O' at 100 km. is greater than that observed by a factor of 10; it is therefore concluded that a core

polarization of this amount is generally untenable for the upper atmosphere. With Yamanouchi's value of $\sigma(T)$, it is seen that at 100 km., 300 km. and 400 km. computed values of ν_0' are of the correct order of magnitude.

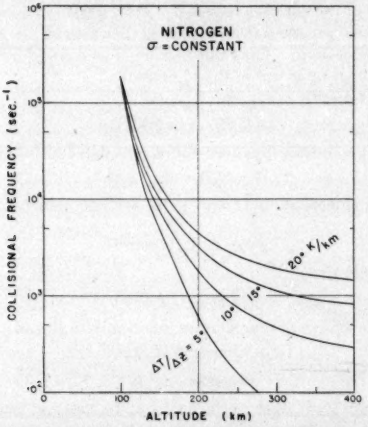


Figure 8. Computed collisional frequencies of electrons with nitrogen particles (both atomic and molecular) as a function of altitude considering various temperature lapse rates.

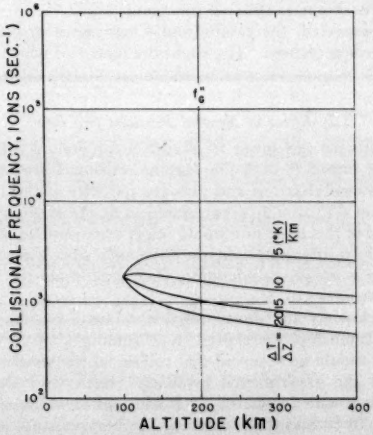


Figure 10. Computed collisional frequencies of electrons and ions for various temperature lapse rates as a function of altitude.

It remains now to consider the effect of electron collisional frequency with ions. For this purpose, equation (4.7) was employed under the conditions of (*a*) different temperature lapse rates and (*b*) an ion density increasing linearly from 2×10^5/cm³ at 100 km. to 4×10^6/cm³ at 400 km. With these assumptions, the curves of Figure 10 were obtained. It is at once clear that at the 100 km. level the effect of electron collisions with ions is negligible in comparison with collisions with neutral particles. It is also obvious that there is an almost constant value of electron–ion collisional frequency for a given temperature lapse rate from about 150 km. to 400 km. or possibly to even greater altitudes. This constant value may mask the lower electron–neutral particle collisional frequencies at great elevations.

From the various results (Figures 8–10) it now is possible to synthesize conditions in the atmosphere. Observed collisional conditions at 100 km. are only obtained when nitrogen particles and oxygen atoms having Yamanouchi's value of $\sigma(T)$ are present at this level. At 300 km. computed collisional frequencies for the combined electron reactions with ions, nitrogen and 'Yamanouchi's oxygen' agree with the experimental values of ν' if $\Delta T/\Delta z$ is 10·5° K/km.; similarly, at 400 km. theoretical results are also very close to observed values if this lapse rate exists.

Table 5. Temperatures deduced from Collisional Frequency and Hypsometry

z (km.)	100	150	200	250	300	350	400
T (° K.)	300	825	1350	1875	2400	2925	3450
T (° K.)*	300			1200			

* From Martyn and Pulley (1936); see also Kidson (1937).

Other computed sets of collisional frequency values do not fit the observed data as well as the combination chosen above. For example, at 100 km. the computed collisional effects of ions together with those of atomic oxygen when $p=5{\cdot}7$ are negligible in comparison with those of nitrogen; yet the evaluated results of electron collisions with nitrogen are smaller than those experimentally determined. The effect of oxygen with $p=18{\cdot}3$ is obviously too large.

The most probable temperature–height relationship based upon discussions of this section is given in Table 5.

5.3. *Comparison with Other Investigators*

Martyn and Pulley (1936) in attempting a derivation similar to (5.5) arrived at the relationship

$$\nu'/\nu_1' = (T_1/T)^{mg/\alpha kT_1}. \qquad \ldots\ldots(5.13)$$

Comparison of (5.4) and (5.13) shows that the term $+\frac{1}{2}$ in the exponent and the factor $(\sigma/\sigma_1)^2$ are missing from the right side of their expression. As a basis for their calculations, Martyn and Pulley used a collisional frequency in the F ionospheric region 1/200 that of the E region, a result in agreement with that given in Table 3. However, they also assumed a mean velocity of electron agitation in the F layer tenfold that in the E region, an assumption which is not fully justified. The results of these authors, obtained by the use of the incorrect equation (5.13) and the conditions given immediately above are listed in Table 5.

§6. TEMPERATURE AND SCALE HEIGHT

6.1. *Scale Heights from Auroral Data*

The definitive expression for scale height is

$$H = kT/mg = kT(R+z)^2/mg_0R^2. \quad \ldots\ldots(6.1)$$

Since k and R are constants, and, in the altitude range under consideration, z at most causes a variation of less than 5%, major changes in the value of H may be attributed directly to changes in either m or T.

At this time it is not intended to evaluate H from considerations of electron concentration in the ionospheric layers; calculations of this type will be discussed in the following section. It is sought here to obtain H by a method independent of the electron concentration. Such a technique has been evolved by Harang (1945, 1946) who obtained values of the scale height from luminosity curves and filter photographs of the northern aurorae. The basic supposition involved in the determination of luminosities from the absorption curve of electrons impinging upon the terrestrial atmosphere is that the luminosity L produced in the height interval dz is proportional to dI/dz where I is the electron beam intensity. This development is in parallel with that usually adopted for ionospheric theory, viz. the problem of determining the ionization produced under the action of a

Table 6. Scale Heights and Corresponding Temperatures as determined from Auroral Data

(Winter 1943–4; Tromsö, Norway)

Maximum altitude (km.)	Scale height (km.)	Temperature (° K.)	Maximum altitude (km.)	Scale height (km.)	Temperature (° K.)
100	8·0	219	185	51	1360
110	8·8	240	200	59·5†	1580
120	10·3	280	250	82·5†	2073
130	12·3	334	270	92	2263
140	15·2	411	300	103*	2455
150	19·7	531	350	119·5*	2704
160	28·3	761			

* Extrapolated. † Interpolated.

monochromatic radiation penetrating a homogeneous atmosphere. Thus an auroral-luminosity function analogous to the ion-production function of Chapman may be derived.

Proceeding in this fashion, Harang obtained the scale heights listed in Table 6 (see Figure 11).

It should be noted that a discontinuity appears in the data at an altitude of about 185 km. Below this level, values of H were computed from data on auroral arcs and draperies, at 185 km. from a study of auroral rays, and at 270 km. from examinations of sunlit rays. Thus the temperatures deduced from data at the 270 km. and higher levels represent daylight conditions. In his work Harang notes that the scale heights deduced represent *minimum* values because the spiral nature of the (incoming) ion orbits was neglected.

Temperatures obtained from observations on the aurorae may differ somewhat from those deduced by other methods. One contributing factor undoubtedly lies in the fact that intensive auroral studies are generally confined to autumn,

winter and spring seasons and to the higher latitudes where temperatures for any given height may be somewhat different from those at lower latitudes. Thus, it should be noted that temperatures deduced from collisional frequencies (which were measured at lower latitudes) are generally higher than those obtained by the use of auroral scale height data. At 350 km., for example, the former method yields temperatures of 3,300° K. while the auroral scale height technique shows

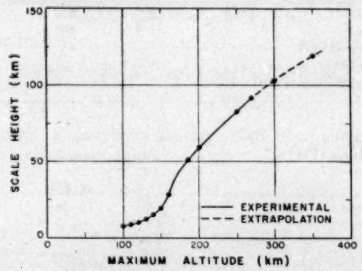


Figure 11. Scale heights as a function of altitude as deduced from auroral data.

only 2,700° K. This difference may arise not only because of the effect of latitude but also because of the neglect of the spiral nature of the orbit. The temperature calculated from auroral scale heights for the northern regions is significantly lower at 100 km. than that deduced by other methods for the same height but at lower latitudes.

6.2. *Scale Heights Determined by Other Methods*

Determination of temperatures from scale heights has been reported previously: Penndorf (1940) performed such calculations based upon the researches of Appleton (1939) for the E region and Pekeris (1940) for the F2 layer. These investigators reported a value of $H = 11{\cdot}4$ km. for the lower and $H = 20$–30 km. for the higher stratum. Using the reported scale heights and different concentrations of molecular oxygen, atomic oxygen, molecular nitrogen and atomic nitrogen Penndorf obtained temperatures in the E region which ranged from 308° to 374° K. For the F2 region, his temperatures ranged from 437° to 625° K. for $H = 20$ km. and from 655° to 936° K. for $H = 30$ km. The small values of H for the F2 region and the corresponding low temperatures obtained therefrom seem unusual.

From ionospheric measurements, Appleton found values of H for the F2 region of 70 km. in summer and 40 km. in winter. Booker and Seaton (1940) found $H = 55$ km. in the autumn and $H = 90$ km. in summer at Watheroo, Australia. Pierce (1947) found a value of $H = 65 \pm 5$ km. during autumn. Kellogg (1950) obtained an average scale height of $H = 170$ km. or the F2 region. Nicolet (1948) gives $H = 30$ km. and $H = 60$ km. for the F1 and F2 regions respectively.

Employing Appleton's values Vassy and Vassy (1942) computed temperatures of 615° K. and 1,080° K. for the F1 and F2 regions. In their calculations they assumed complete nitrogen dissociation. Recomputing these values on the

basis of nitrogen dissociation as given in the model atmosphere we find $T = 955°$ K. and 1,670° K. for the F1 and F2 regions respectively. The above and other reported values of scale height are summarized in Table 7.

Table 7. Values of Scale Height (in km.) determined from Ionospheric Data

Investigator	Ionospheric Region				
	D	E	F*	F1*	F2*
Appleton (1939)		11·4	40		70
Appleton and Beynon (1947)			45	55	45–70
Booker and Seaton (1940)					50–90
Budden, Ratcliffe and Wilkes (1939)	6				
Fuchs (1936 b)					110†
Gerson ‡					67–167
					105–186†
Grace, Kelso and Miller (1949)		10			
Guha (1949)		10			46–54†
Jones (1950)			6–32§		
Jones and Jones (1950)			15–56§		
Kellogg (1950)				43†	170†
Nicolet (1948)				30	60
Pekeris (1940)					20–30
Pierce (1947)					65 ± 5
Rydbeck (1942)					37–58
					51–165†
Rydbeck (1944)			35		

* Region F1 is attributed to a splitting of Region F during daylight. With the approach of darkness, the F1 region rises and coalesces with the F2 region to form a single F region at night. † Equatorial values. ‡ See § 7. § Arctic values.

§ 7. TEMPERATURE AND ELECTRON CONCENTRATION

7.1. *General*

The electron concentration in the ionosphere as a function of altitude for a Chapman distribution is, by (2.14),

$$n = n_{\max} e^{[1-(T)]/2} = n'_{\max} e^{F(T)/2} \qquad \ldots\ldots(7.1)$$

where n is evaluated from

$$n = \pi m f^2/e^2 \qquad \ldots\ldots(7.2)$$

when the Lorentz polarization term is neglected. During ionospheric probings, however, radio transmission frequency is usually measured as a function of virtual height; it is desirable therefore to determine a relationship between the latter variables.

7.2. *Simplifications to Chapman Layer*

Neglecting the terrestrial magnetic field, the virtual height of reflection of the ordinary ray is

$$z' = z_0 + \int_0^s ds/\mu. \qquad \ldots\ldots(7.3)$$

The refractive index μ for the modified Chapman distribution defined in (2.16) is given by (cf. Försterling and Lassen 1931, 1933, Appleton 1937)

$$\mu^2 = 1 - 2s(f_c/f)^2/A + (f_c/f)^2(s/A)^2. \qquad \ldots\ldots(7.4)$$

In the above equations z_0 is the lower level of the layer corresponding to $n=0$, s is measured along z so that $s=0$ at the height $z=z_0$ and $A=z_{max}-z_0$. Integration and simplification yields

$$z'=z_0+(\boldsymbol{k}T/m\boldsymbol{g})(f/f_c)\ln[(f_c+f)/(f_c-f)] \qquad \ldots\ldots(7.5)$$

when $A=2H$.

On the basis of a cubic approximation to a Chapman region, viz.

$$n=n_{max}[1-(32/81)Z^2-(64/729)Z^3] \qquad \ldots\ldots(7.6)$$

Fuchs (1936 b) obtained

$$z'=z_0+(9/16)(\boldsymbol{k}T/m\boldsymbol{g})B\ln[(1+B)/(1-B)] \qquad \ldots\ldots(7.7)$$

when $A=9H/4$ and where $2B^2=-1+(1+8f^2/f_c^2)^{1/2}$. $\ldots\ldots(7.8)$

Figure 12 shows a graph indicating both the cubic and parabolic approximations to a Chapman region.

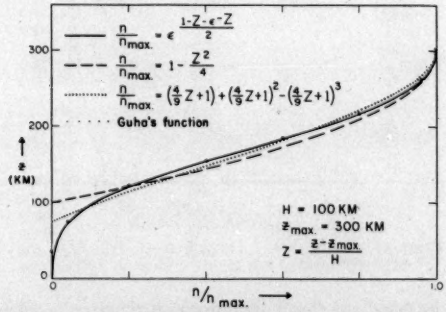


Figure 12. Graph of the Chapman distribution, and its cubic, quadratic and Guha approximations.

To emphasize the similarity between them, equations (7.5) and (7.7) may be written in the general form

$$T=(m\boldsymbol{g}/\boldsymbol{k})(z'-z_0)/\phi(f) \qquad \ldots\ldots(7.9)$$

where in Appleton's formula

$$\phi(f)_A=(f/f_c)\ln[(1+f/f_c)/(1-f/f_c)] \qquad \ldots\ldots(7.10)$$

and in Fuchs' formula

$$\phi(f)_F=(9B/16)\ln[(1+B)/(1-B)]. \qquad \ldots\ldots(7.11)$$

In both cases, it should be noted that the influence of the geomagnetic field has been neglected. A graph of $\phi(f)_A$ and $\phi(f)_F$ as a function of f/f_c is shown in Figure 13.

On the basis of (7.9) with $\phi(f)=\phi(f)_A$ Appleton obtained a value $(\boldsymbol{k}T/m\boldsymbol{g})=70$ km. for the F2 ionospheric layer in winter. With $\phi(f)=\phi(f)_F$ the magnitude of $(\boldsymbol{k}T/m\boldsymbol{g})$ would be greater, the exact value depending upon the radio frequencies involved.

The simultaneous solution of equation (7.9) for conditions at two successive virtual height levels z_1' and z_{1+1}' gives

$$T=(mg/k)[\Delta z'/\Delta\phi(f)] \qquad \ldots\ldots(7.12)$$

where $\Delta z'=z_1'-z_{1+1}'$ and $\Delta\phi(f)=\phi(f_1)-\phi(f_{1+1})$.

Using (7.12), Fuchs evaluated temperatures for the F1 and F2 regions over Huancayo, Peru, using a value of $mg/k=1{\cdot}77\times10^{-4}\,°\text{K/cm}$. This value appears somewhat small and indicates almost complete nitrogen dissociation. Inasmuch as experimental evidence does not verify such large degrees of dissociation, a better value of mg/k based upon the constants given in this work may be used. Thus, for example, at 185 km., $mg/k\simeq2{\cdot}67\times10^{-4}\,°\text{K/cm}$. with the value of nitrogen dissociation adopted in the model atmosphere.

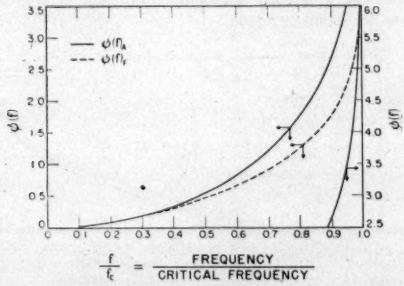


Figure 13. Graph of $\phi(f)_A=(f/f_c)\ln[(1+f/f_c)/(1-f/f_c)]$ and $\phi(f)_F=(9B/16)\ln[(1+B)/(1-B)]$ where $B=B(f/f_c)$.

It should be noted also that a discrepancy, attributable to rounding errors, appears in Fuchs' paper. At one stage of his computations, two significant figures were employed. As subsequent calculations involved a subtraction and division, an error of one unit introduced considerable error in Fuchs' evaluated temperatures. A comparison of Fuchs' values and those evaluated with greater computational accuracy and a better value for mg/k is given in Table 8. These values are for the F1 and F2 layers in February, 1934, at noon.

A third relationship for the evaluation of temperatures by approximations to the Chapman region has been obtained by Guha (1949). As an approximation to the Chapman ion distribution, Guha writes

$$-Z=0{\cdot}6672+1{\cdot}397(n/n'_{\max})-2{\cdot}084(n/n'_{\max})^{1/2}+2(1-n/n'_{\max})^{1/2}. \qquad \ldots\ldots(7.13$$

Use is also made of the general expression for the refractive index of an ionized gas in the absence of an external magnetic field

$$\mu^2=1-ne^2/\pi mf^2. \qquad \ldots\ldots(7.14)$$

Substituting (7.13) in (7.3) and integrating in terms of μ, we obtain immediately

$$z'=z_{\max}-2{\cdot}667H-2{\cdot}794H(f/f_c)^2+3{\cdot}274H(f/f_c)$$
$$+2H(f/f_c)\sinh^{-1}[(f_c/f)^2-1]^{1/2} \qquad \ldots\ldots(7.15)$$

when the refractive index is written in the form

$$\mu^2 = 1 - (f_c/f)^2(n/n'_{max}). \qquad \ldots\ldots(7.16)$$

Equation (7.15) may also be solved for z_{max} and H as indicated in (7.12) using data taken from two points of the experimental virtual height–frequency curve.

A fourth method also exists for obtaining temperature by a procedure similar to those outlined above. In most cases of interest a probing electromagnetic wave

Table 8. Temperatures at Huancayo, Peru, at Noon L.M.T. in February 1934

	Frequency (Mc/s.)	Temperature (° K.) Fuchs	Recomputed using $\phi(f)_E$	Recomputed using $\phi(f)_A$	Altitudes (km.) Fuchs	Recomputed
F1 layer 14.2.34	3·53	373	553	316	190–220	185
	4·00	368	599	333		
	4·21	782	1184	852		
	4·37					
	4·52					
F1 layer 28.2.34	3·50	435	623	355	190–220	185
	3·90	460	737	410		
	4·11	1040	1626	897		
	4·23					
	4·41					
F2 layer 14.2.34	6·30	1670	1942	1113	350–420	355
	6·60	1510	2086	1170		
	6·90	1960	2448	1359		
	7·10	1960	2384	1324		
	7·26					
	7·75					
F2 layer 28.2.34	5·45	1175	1442	826	350–420	395
	5·80	980	1198	672		
	6·20	1430	1717	947		
	6·39	1950	2403	1332		
	6·49					
	6·80					

begins and terminates its ionospheric path where the refractive index is closely unity; thus it is possible to determine the excess of equivalent path over the actual distance travelled. This excess is, from (7.3),

$$X = \int \left(\frac{1}{\mu} - 1\right) ds \qquad \ldots\ldots(7.17)$$

(cf. Jaeger 1947). Define

$$x^2 = e^{-Z}/2 \quad \ldots\ldots(7.18); \qquad a = (2e)^{1/2}(f_c/f)^2. \quad \ldots\ldots(7.19)$$

The refractive index given in (7.14) then becomes

$$\mu^2 = 1 - ax\exp(-x^2) \qquad \ldots\ldots(7.20)$$

and equation (7.17) is

$$X = -2H\int_{x_1}^{\infty} [1/\{1 - ax\exp(-x^2)\}^{1/2} - 1](dx/x). \qquad \ldots\ldots(7.21)$$

Using the greater root x_1 of $ax = \exp(x^2)$ .....(7.22)

the height of reflection above the earth's surface may be obtained from

$$z = z_{\max} + H \ln(\tfrac{1}{2}) - 2H \ln x_1. \qquad \text{.....(7.23)}$$

The virtual height of reflection is thus obtained from

$$z' = z_{\max} + H \ln(\tfrac{1}{2}) - 2H \ln x_1 + HY(f/f_c)/2. \qquad \text{.....(7.24)}$$

Values of $x_1 = x_1(f/f_c)$ and $Y = Y(f/f_c)$ are tabulated by Jaeger (1947).

Equations (7.5), (7.7), (7.15) and (7.24) are valid only if no lower ionospheric regions are present. If such a lower retarding layer is present, the following formulae apply

(*a*) In place of (7.5),

$$z' = z_0 + H_F(f/f_c) \ln[(f_c+f)/(f_c-f)] - 4H_E + 2H_E(f/f_E) \ln[(f+f_E)/(f-f_E)]. \qquad \text{.....(7.25)}$$

(*b*) In place of (7.15)

$$z' = z_{\max} + H_F\{-2{\cdot}667 - 2{\cdot}794(f/f_c)^2 + 3{\cdot}274(f/f_c) + 2(f/f_c)\sinh^{-1}[(f_c/f)^2 - 1]^{1/2}\} + AH_E. \qquad \text{.....(7.26)}$$

(*c*) In place of (7.24)

$$z' = z_{\max} + H_F[\ln \tfrac{1}{2} - 2 \ln x_1 + Y(f/f_c)/2] + H_E Y'(f_E/f)/2. \qquad \text{.....(7.27)}$$

In the latter group of equations H_F and H_E refer to the scale heights of the F and E layers respectively, f_c is the critical frequency of F layer, f_E the critical frequency of E layer and

$$A \equiv (\pi e/2)^{1/2}(f_c/f)^2 + (3e/8)(f_c/f)^4 + (5/24)(\pi e^3/6)^{1/2}(f_c/f)^6 + \ldots \qquad \text{.....(7.28)}$$

where e is the base of natural logarithms.

Another expression for the evaluation of temperature may be derived from (2.16):

$$T = mg(z - z_{\max})/2k(1 - n/n'_{\max})^{1/2}. \qquad \text{.....(7.29)}$$

On this basis Pekeris (1940) obtained the rather low values $H = 20$–30 km. for the F2 region. The same data and equation (7.29) yielded $H = 18$–35 km. while computations from (7.5) indicate $H = 28$–32 km. Thus Pekeris' work is checked and it is concluded that on the dates in question the scale height over England was somewhat small.

That the various techniques described above are approximations is clear in most cases. Jaeger's procedure rests upon a numerical approximation to the integral involved. Appleton and Fuchs both obtained analytic solutions for z' by assuming that the electron density is approximated by appropriate polynomials in z.

7.3. *Assumption of a Chapman Layer*

Solutions may be obtained for T from the complete expressions for a Chapman region (equations (7.1) and (7.2)):

$$4 \ln(f/f_c) = F(T). \qquad \text{.....(7.30)}$$

This equation was used to determine F2 region temperatures at Washington, D.C., and Huancayo, Peru, using the appropriate data for noon on 14th January 1947. For purposes of comparison, temperatures at the same stations were also determined from several of the preceding equations. True heights of maximum ionization, obtained by the method originally given by Appleton (1930) and recently restated by Manning (1947) are compared with those computed from some of the above relationships. Results are tabulated in Table 9.

Table 9. Comparison of Results: Ionospheric Region Data
Local Noon 14th January 1947

Method	Equation No.	z_{max}(km.)*	H_F(km.)*	H_E (km.)	T(° K.)*
Washington, D.C.					
de Groot†	—	338	—	—	—
Chapman	7.5	331	67	—	1530
Chapman	7.30	—	71	—	1630
Chapman	7.25	186	68	−550 to +440	—
Fuchs	7.7	—	117	—	2680
Guha	7.15	313	68	—	1600
Guha	7.26	320	81	−305 to +142	—
Jaeger	7.24	309	73	—	1720
Jaeger	7.27	284	83	−580 to +110	—
Huancayo, Peru					
de Groot†	—	484	—	—	—
Chapman	7.5	500	105	—	2200
Chapman	7.30	—	132	—	2750
Fuchs	7.6	—	186	—	3910
Guha	7.15	564	160	—	—
Jaeger	7.24	568	154	—	—

* F2 region. † See Appleton (1930), Manning (1947).

The temperature may also be estimated from the quasi-Chapman equation (2.5)

$$Q = \beta n'' J_\infty \rho_i (T/T_i)^{-(1+A)} \exp -[\rho_i \beta H_i \sec \chi (T/T_i)^{-A}] \quad \ldots\ldots (7.31)$$

which takes account of the effect of a linear temperature lapse rate. From (7.31) we may obtain

$$z_{max} = z_i + (1/b)\{[(\cos\chi/\rho_i\beta H_i)(1+A)]^{-A} - 1\} \quad \ldots\ldots (7.32)$$

and $$n_{max} = (\beta\rho_i n'' J/\alpha)^{1/2}[(\cos\chi/\rho_i\beta H_i)(1+A)\exp\{-(1+A)/2\}] \quad \ldots\ldots (7.33)$$

where $A = 1/bH_i$. The above equations apply only for a linear temperature increase with altitude, and the subscript i refers to the value of the quantity concerned at a datum level.

Using ionospheric data obtained at South Africa during the period November 1945 to February 1946 Gledhill and Szendrei (1950) estimated temperature values at 200 km. and 300 km. on the basis of equations (7.31) to (7.33). Assuming $M = 15$ (so that $mg/k = 16{\cdot}2°$ K/km.) and $\rho_i = 10^{-11}$/cm³ at a datum level $z_i = 200$ km. solution was made for b, T_i and β using appropriate values of χ. To determine probable values of b and T_i, consistency examinations were made of the computed values of absorption coefficient β and the factor $(\beta\rho_i n'' J/\alpha)$.

Upon the adoption of $\beta = 3{\cdot}2 \times 10^{17}\,\mathrm{cm}^2$, it was concluded that probable 200 km. and 300 km. temperatures (in °K.), as obtained from the above data, are:

Altitude	Probable Temperatures (°K.) by Months			
(km)	Nov.	Dec.	Jan.	Feb.
200	485	408	369	447
300	1097	1376	1392	1305

7.4. *Conclusions*

The results given in Table 9 reveal the care which must be taken in applying some of the previously developed equations. Confining attention to results obtained when the effect of retardation by a lower ionospheric layer is neglected (equations (7.5), (7.7), (7.15) and (7.24)), the computed values of temperature seem to indicate a generally higher temperature in the tropics than in middle latitudes. Thus at Washington values range from 1,530 to 2,680° K. but at Huancayo the range is from 2,200 to 3,910° K. The spread in computed temperature values is certainly considerable in each instance, especially since the same data were originally used.

In examining the cause for the internal discrepancies of the evaluated results, it should be noted that the equation (7.1), basic to all the above computations, is in reality inapplicable to the F2 region, for this layer can scarcely be called a Chapman region which, incidentally, implicitly assumes an isothermal atmosphere.

Approximations to the original (incorrect) hypothesis thus not only retain the original errors but also those introduced by the approximations. It should also be recalled that (7.1) was obtained on the assumption that $dn/dt = 0$; this assumption is usually not true for the F regions.

The obtaining of experimental values of critical frequency and virtual height also present some difficulties. On some occasions the true electron density near the level of maximum ionization cannot be accurately determined; also, near the critical frequency, the estimation of the virtual height is not easily obtained and is thus subject to error.

An additional factor which may affect the computed results appreciably is the neglect of the earth's magnetic field. Preliminary calculations reveal that neglect of the geomagnetic field causes a change in the scale height of the E region by a factor of about 1·2. An error of even greater magnitude may be introduced in the determination of H for the F2 region. Atmospheric tidal oscillations have been shown to influence the value of scale height (Jones and Jones 1950); the effect of such tides also has not been considered.

When the effect of retardation of a probing electromagnetic wave by a lower ionospheric layer is considered (equations (7.25), (7.26) and (7.27)), the computed values of scale heights for the E region are in gross disagreement with expected results. The obviously incorrect values may be attributed not only to the various factors mentioned above but also to the fact that each equation was obtained on the assumption that the lower retarding layer was parabolic or Chapman in form. Actually, the retardation is also markedly influenced by the character of the electron distribution *between* the layers; this distribution may entirely vitiate the usefulness of the equations. Also the solution is very sensitive to variation in virtual height, so that a small error made in determining the virtual height

seriously affects the calculated values of the E-region scale height H_E. Evaluations of H_E shown in Table 9 cast serious doubt on the usefulness of these equations.

While the quasi-Chapman formula (7.31) is already on a firmer basis for the determination of temperatures, it also suffers from some of the deficiencies stated above: thus, this equation was derived on the basis that $dn/dt=0$ and that no magnetic field exists. The same difficulties in obtaining accurate values of electron density and critical frequency also exist. In equations (7.31) to (7.33), a further complication arises from the lack of an accurate value of the absorption coefficient β to be used for the computations. Also, the value of M used in the evaluations undertaken by Gledhill and Szendrei indicates almost complete nitrogen dissociation. While this latter method, which assumes a linear temperature gradient, is undoubtedly superior to the assumption of an isothermal region, it still cannot be directly applied to the F2 region because of the many peculiarities of this layer.

On the basis of the preceding results, it must be concluded with Pande (1947) that the application of experimental values of electron concentration to the determination of ionospheric temperatures in the F2 layer at least is open to considerable question. The techniques give results of the proper order of magnitude and thus a rough estimation of temperature. Undoubtedly this method has been overworked.

§8. TEMPERATURE AND THE RECOMBINATION COEFFICIENT

8.1. *Recombination at Low Pressure*

Before attempting to consider this problem in detail a clear understanding must be obtained of the action implied by the term recombination. Recombination of ions occurs when a positive and negative ion (including electron) associate so as to form a neutral system. The coefficient of recombination defines the rate at which neutral systems are formed.

According to this definition recombination implies the consummation of a process which occurs under a variety of circumstances. The methods involved in each case may differ markedly. Reasons for such variations lie in the great differences in the nature of the ion carriers and the states in which they remain after ionization. For example, the mechanism whereby a free electron combines with a positive ion differs radically from that involved when a charge transfers from a normal negative to a normal positive gaseous ion, or when two oppositely charged ions neutralize by contact. While each action is a recombination, the mechanism and the characteristic constants in each instance are distinct and require separate consideration.

The general term recombination includes such specialized actions as electron, preferential, columnar, and volume recombination, each describing a specific mechanism having its appropriate recombination coefficient. Furthermore, in ionospheric studies Appleton and Sayers (1938), Ghosh (1944), Massey (1937) and others have defined a generalized recombination coefficient which describes the net effect of coexistent attachment and recombination, or the net effect of several simultaneous or successive recombination mechanisms. Considerable confusion may exist whenever the term recombination coefficient is loosely employed.

Fortunately, in ionospheric work only two types of mechanisms need be considered, (*a*) volume recombination at low pressures and (*b*) generalized recombination. The former reaction occurs when ions of opposite sign are found in equal numbers randomly distributed in space—a condition found in the F ionospheric region. The latter mechanism is generally considered to exist in the E layer.

Volume recombination at low pressures may be defined by

$$\alpha = \pi(\overline{u^2} + \overline{v^2})^{1/2}(Rd^2 + \sigma^2/4) \qquad \ldots\ldots(8.1)$$

where

$$d = 2e^2/3kT = \text{radius of active attraction} \qquad \ldots\ldots(8.2)$$

u, v = velocity of each type of ion respectively and R is the probability of impacts within the distance d. The term σ may be considered as the average effective collisional diameter under the assumption of a Maxwellian velocity distribution.

Depending on whether the energy of the ions at the last encounter is less than, equal to, or greater than that gained in a free fall from infinity under the attractive forces, the orbits will be closed ellipses, parabolae or hyperbolae. If the value of d is of the order of a free path or less, the ions will usually approach each other with energies equal to the sum of the thermal energy and the energy gained in their mutual potential field. Thus the orbits will be open and the ions have a good probability of separating after passing each other. If, however, they make kinetic impacts with neutral molecules within the distance d they have a fair chance of losing part of the energy gained in the field, and therefore of being captured into elliptical orbits. Thomson and Thomson (1928) postulated that an impact of one of the ions with a third particle within the distance d would generally serve to reduce the energy sufficiently to close the orbit. Thomson's assumption will be adopted here.

At low pressures the mean free path between collisions is large and $R \simeq (4d/3)(1\ \bar{\lambda}_1 + 1/\bar{\lambda}_2)$ where the subscripts refer to ions of the first and second type respectively. Since $\sigma/2 \simeq 10^{-8}$ cm. and $d \simeq 10^{-7}$ cm., the term Rd^2 is the dominant term at pressures greater than about 10^{-5} mb., whereas the term $(\sigma/2)^2$ is the dominant term at lower pressures. Thus, when the pressure is greater than 10^{-5} mb. (8.1) may be effectively simplified to

$$\alpha = \pi R d^2(\overline{u^2} + \overline{v^2})^{1/2} \qquad \ldots\ldots(8.3)$$

which is the equation obtained from the Thomson theory.

At lower pressures, as its first term may be neglected (8.1) becomes the relationship given by the very low pressures theory (Loeb 1939)

$$\alpha = \pi(\sigma/2)^2(\overline{u^2} + \overline{v^2})^{1/2}. \qquad \ldots\ldots(8.4)$$

As $\bar{\lambda}$ varies as $T^{1/2}$, d as T^{-1} and $(\overline{u^2} + \overline{v^2})^{1/2}$ as $T^{1/2}$, equations (8.3) and (8.4) give, respectively

$$\alpha = AT^{-3} \qquad \ldots\ldots(8.5), \qquad \alpha = BT^{1/2} \qquad \ldots\ldots(8.6)$$

with the transition from the former mechanism to the latter at a pressure of the order of magnitude of 10^{-5} mb. The coefficients A and B are dimensional constants.

The physical reasoning behind (8.1) is readily apparent. The term Rd^2 describes the mechanism whereby one ion follows a closed orbit around a second

ion after colliding with a third body within the distance d. As the pressure is reduced, however, the probability R of such an action decreases until the magnitude of Rd^2 equals that of $(\sigma/2)^2$. Under these circumstances the probabilities of a direct neutralization collision and the three-body Thomson process are about equal. Any further reduction in pressure rapidly reduces the former term in comparison with the latter. As pressures in the F2 region are generally considered to be lower than 10^{-6} mb., and attachment is negligible, condition (8.6) rather than (8.5) should be used for this layer. The applicability of (8.5) to the E layer will be discussed below.

It should be noted that (8.5) and (8.6) may be considered as special equations of a general family

$$\alpha = CT^r. \qquad \ldots\ldots(8.7)$$

Inasmuch as $\alpha_a = \alpha(T_a)$, it is possible to write

$$\alpha/\alpha_a = (T/T_a)^r. \qquad \ldots\ldots(8.8)$$

Milne (1924), in his theory on statistical equilibrium, as well as Yamanouchi and Kotani (1940), obtained the relationship $r = -1/2$ when T signifies the electron temperature T_e. Tukada (1937) obtained a similar result based on quantum mechanical considerations. From the results of a theoretical study by Bates and collaborators (1939) of low-pressure, radiative recombination of electrons with positive ions, Gerson (1949 a) obtained a value $r = -3/4$. Bates and Massey (1946) indicate that $r = -3/2$ under ionospheric conditions. Thomson's (1924) early theoretical work led to values $r = -5/2$ at low pressures and $r = -3/2$ at high pressures. Biondi and Brown (1949) arrived at a theoretical value of $r = -0{\cdot}9$ based upon quantum mechanical considerations.

This variety of values must be interpreted in the light of the recombination mechanisms involved. For radiative recombination of electrons with positive ions the probable value of r is $r = -1/2$, and for electronic volume recombination of the type found in discharge tube plasma, probably $r = -3/2$. However, these reactions do not seem applicable to the ionosphere at the present state of knowledge.

In the E region the processes of attachment and recombination are considered to be coexistent. The combined action may be described by a generalized or effective coefficient of recombination (Sayers 1943)

$$\alpha' = \alpha_e + K\alpha_i \qquad \ldots\ldots(8.9)$$

where α_e is the electron–positive ion recombination coefficient, α_i the negative ion–positive ion recombination coefficient, $K = n_-/n_e$ and n_- is the number of negative ions.

Although α_e and α_i may be functions of temperature of the type described by (8.7), there is no reason to assume that in each case the values of r are identical. Under these circumstances, the direct application of (8.5) to the E ionospheric layer does not appear advisable.

In several comprehensive papers, Seaton (1947 a, b, 1948) undertook an intensive and extensive evaluation of upper atmospheric temperature distributions from 100 to 300 km., including the course of temperature throughout the day at given latitudes. Before evaluating the temperature by this method, however, it was first necessary to determine the recombination coefficient throughout the day.

8.2. *Diurnal Cycle of Recombination Coefficient*

The recombination coefficient at various times during the day has been obtained as follows:

(*a*) During darkness $dn_a/dt = q_a - \alpha_a n_a^2$, where $q_a = 0$. .....(8.10)

(*b*) For comparatively short but not necessarily equal intervals on either side of sunrise or sunset

$$dn_a/dt = q_a - \alpha_a n_a^2;\ dn_b/dt = q_b - \alpha_a n_a^2 \qquad .....(8.11)$$

i.e. assuming $\alpha_b = \alpha_a$ and $n_b = n_a$, and again $q_a = 0$.

(*c*) For comparatively short and equal intervals on either side of noon

$$dn_c/dt = q_c - \alpha_c n_c^2;\ dn_d/dt = q_c - \alpha_c n_d^2 \qquad .....(8.12)$$

i.e. assuming $q_c = q_d$ and $\alpha_c = \alpha_d$.
In each case, solution may be made for α. It is apparent that with symmetrical conditions on either side of noon, $n_c^2 = n_d^2$, and equations (8.12) become inconsistent.

It should be noted that (8.11) and (8.12) may be valid for the F2 region only when the level of maximum ionization has not been disturbed by an alteration in atmospheric conditions but remains at the same height during the period considered. The validity of these equations for *extended* periods on either side of sunrise (or sunset) and noon respectively is open to serious question. Preliminary studies by Seaton (1947 c) seemed to indicate somewhat reasonable and consistent results when the intervals on either side of noon were taken as of the order of several hours. Such conclusions, however, do not seem warranted. *Inasmuch as large discrepancies may arise from these procedures it may be safely stated that with very few exceptions the use of* (8.10) *to* (8.12) *to determine* α *and* q *for overlapping time intervals is unjustified.*

In his evaluations, Seaton used for the E, F1 and F2 ionospheric layers the relationship

$$\alpha/\alpha_a = (T/230)^{-3}. \qquad .....(8.13)$$

The value $T = 230°$ K. was taken from that determined by Vegard and Tönsberg (1935, 1938) from band spectra of the aurora borealis. The adopted value of α_a was chosen as follows: (*a*) for the E and F1 ionospheric regions, the earliest morning value of α; (*b*) for the F2 layer, the mean of the nocturnal values of α calculated from (8.10). (In the first paper (Seaton 1947 b) the value α_a chosen for use at three stations was the mean 24-hour recombination coefficient for Tromsö, Norway.) On this basis temperature throughout the 24-hour period for 18 stations ranging in latitude from 70° N. to 43° S. and fairly well distributed longitudinally were calculated for January 1947. For the E and F1 regions during daylight, temperatures were found to vary from 71° K. to 568° K. and from 35° K. to 1,785° K. respectively. In the F2 layer temperatures ranged from 84° K. to 1,098° K. over the 24-hour period.

8.3. *Limitations of the Method*

Several limitations inherent in any method of deducing ionospheric temperatures from recombination coefficients should be noted. Such limitations arise from (*a*) techniques employed for obtaining the recombination coefficient, and (*b*) uncertainties regarding the corresponding values of T and α to be used.

In the determination of α throughout the day from equations (8.10) to (8.12) it is obvious that as the slope of the (n, t) curve approaches a constant near noon $\alpha \to 0$.* Thus $T \to \infty$ if $r < 0$ or $T \to 0$ if $r > 0$. Furthermore, there may be occasions during the evaluations when apparently $\alpha < 0$. The occurrence of these curious results, which cast doubt upon the methods used, may be traced to the basic assumptions underlying (8.11) or (8.12). The validity of the former equation, for example, to obtain α for extended periods on either side of the departure point, noon, has not been conclusively demonstrated. Further, the utility of these two equations depends upon (*a*) changes in ionic concentration being represented by the quadratic law of recombination, and (*b*) no disturbances taking place in the level of maximum ion density. Obviously (see § 9) neither assumption is valid if large temperature variations occur during a 24-hour period since the resulting influx or efflux of ions has been neglected.

If the method of obtaining the recombination coefficient during the 24-hour period is accepted, other problems arise in connection with the determination of the corresponding values of α and T. Early laboratory experiments by Erikson (1909, 1912) have yielded a value of $r \simeq -7/3$ and by Phillips (1910) of $r = -2{\cdot}2$ (both somewhat close to $r = -3$), at pressures far in excess of those found in the ionosphere, and presumably for a truly recombinational process. The extrapolation of such a value, even when considering its general agreement with theory, for use with an effective recombination coefficient hardly seems warranted.

Finally there is the question of corresponding values of T and α. Since $\alpha = \alpha(T)$, any given value of the recombination coefficient has a corresponding value of temperature, and it would seem that this value of temperature should be employed in the equations rather than that of rotational temperature evaluated from auroral observations. Indeed, it has been shown by Ornstein and Haringhuizen (1932) and Horie (1940) that temperatures obtained from band rotational spectra are not necessarily equal to the gas temperatures. Duffendack and collaborators (1934, 1939) on investigating the temperature deduced from nitrogen negative bands concluded that rotational temperatures depend not only upon the gas temperature, but also upon the type of impact, i.e. electron impact, collisions of the second type, and undoubtedly others. Similarly Thomson (1934, 1935) noted that the effective rotation temperature is not identical with the gas temperature although the former is dependent upon the latter. Thomson reported that rotational temperatures derived from the N_2^+ bands were lower than the temperature of the furnace surrounding the discharge tube.† Branscomb (1950) found abnormally low rotation in molecules excited by electron impact—a condition which may easily exist during auroral displays.

Reference values of α and T may be computed from (8.4) for the F2 ionospheric region. Considering $\sigma = 3{\cdot}2 \times 10^{-8}$ cm., and $\overline{u^2} = \overline{v^2} = 3\boldsymbol{k}T/m$ (m = mass of oxygen atom or ion) and $T = 2{,}000°$ K., it is found that $\alpha = 2 \times 10^{-10}$ cm^3/sec. This value is remarkably close to that obtained by Wells and Shapley (1946), Sayers (1943) and others in evaluating α from ionospheric conditions.

* The electron density in the F2 layer may increase through the noon period. If the rate of increase is linear, $dn_b/dt = dn_b/dt$ and α becomes 0.

† Temperatures deduced from rotational band spectra may be materially lower than the gas temperatures under somewhat specialized conditions. These conditions may be found in the ionosphere and mesosphere. See Schüler and Gollnow (1938, 1939).

In Seaton's latest paper (1948) the recombination coefficient for the F2 ionospheric region varies from 1×10^{-12} to $1{\cdot}3\times10^{-9}$ with corresponding temperatures (when $r=-3$) varying from about 1,100° K. to about 85° K. respectively. These computations were based upon the assumption that the mean night-time value of α corresponded to the temperature $T=230°$ K. Had the value $r=+1/2$ been chosen with the values $\alpha=2\times10^{-10}$ at $T=2{,}000°$ K., the corresponding temperature range would have been 5×10^{-2} °K. to 84,500 °K.! Truly the method of calculating temperature from recombination coefficients requires considerably more study.

Because of the objections given above, applications of this method for correlating surface and F2 layer ionospheric temperatures, as given by Baranul'ko (1948), are open to serious criticism.

8.4. *Seasonal F2 Region Temperature Variation*

The relationship between temperature and recombination coefficient has also been used to glean information regarding a possible seasonal or latitudinal temperature gradient. Thus from (2.17), assuming $J_a=J_b$, $m_a=m_b$ and $\boldsymbol{g}_a=\boldsymbol{g}_b$

$$n_a/n_b=[(T_b\alpha_b\cos\chi_a)/(T_a\alpha_a\cos\chi_b)]^{1/2} \qquad \text{.....(8.14)}$$

using (8.8)
$$n_a/n_b=(T_b/T_a)^{(r+1)/2}(\cos\chi_a/\cos\chi_b)^{1/2}. \qquad \text{.....(8.15)}$$

If conditions a and b represent opposite solstitial periods at noon at the same geometric altitude above the surface of the earth,

$$n_a/n_b=(T_b/T_a)^{(r+1)/2}[\sin(\theta+\delta_a)/\sin(\theta+\delta_b)]^{1/2} \qquad \text{.....(8.16)}$$

where the subscript a is now identified with December and b with June. Employing middle latitude values, for example, $\theta=45°$ N. or 45° S., $\delta_a=-\delta_b=-23{\cdot}5°$ and making the reasonable assumption, as confirmed generally by observation, that $n_a=2n_b$, June temperatures in both the northern and southern hemispheres expressed in terms of December temperatures for various values of r are shown in Figure 14.

As would be anticipated, the diagram indicates a pole at $r=-1$. When $r<-1$ temperatures at latitude 45° in both hemispheres are lower during June than during December. Thus, if $r=-3$, June temperatures at 45° N. are about 0·25 the value of December temperatures, and at 45° S. about 0·8 the value of December temperatures. Such a result seems curious for the northern hemisphere at least. Of course, during December the sun is closer to the earth than during June, but it seems doubtful that the closer proximity overbalances the effect of the decreased angle of incidence in the northern hemisphere. An equally curious result is obtained when $r>-1$, for in this instance noon temperatures in June are greater than those during December in both hemispheres. It may be observed when $r=0{\cdot}5$, the ratio $T_{\text{June}}/T_{\text{Dec.}}$ is about 6 in the northern hemisphere at a latitude of 45° N. and about 1·3 in the southern at a latitude of 45° S.

The determination of seasonal variations of temperature in the ionosphere has been attempted in the past. Appleton and Naismith (1935) used (8.14) and derived the equation

$$n_a/n_b=(T_b/T_a)^{1/2}[\sin(\theta+\delta_a)/\sin(\theta-\delta_a)]^{1/2}, \qquad \text{.....(8.17)}$$

i.e. $r=0$.

This relationship, obtained on the assumption that the electron temperature during winter equalled that during summer, essentially states that the recombination coefficient was constant throughout both periods. It had been found that electron densities during December were higher than those occurring in June in Great Britain, Japan and the United States, and from that fact arose the thermal expansion hypothesis of Appleton and Naismith. Under their theory greater summer temperatures in the F2 region caused the entire layer to expand vertically, thereby decreasing the normal electron concentration. During winter the opposite effect was assumed operative, i.e. a temperature decrease produced a contraction, thus increasing the ionic concentration in this region. Midsummer noon temperatures in the F2 ionospheric region were computed

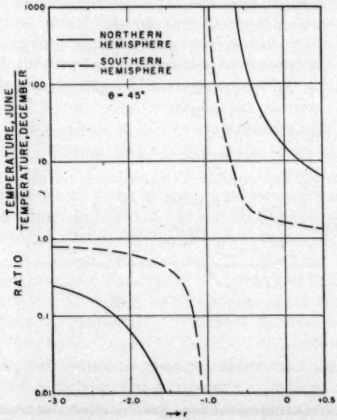


Figure 14. Ratio of June to December temperatures expressed as a function of r. For the northern hemisphere the curves represent conditions at latitudes 45° N.; for the southern hemisphere, conditions at latitudes 45° S.

to be at least four times those occurring at midwinter noon. By arbitrarily adopting a value of 300° K. for midwinter darkness temperatures, Appleton and Naismith then concluded, on the basis of F2 ionospheric electron densities observed over Great Britain, that summer temperatures are at least 1,200° K.

If $r = -0{\cdot}5$, the same authors note that the ratio of summer to winter temperatures may be about 10 to 1 at the latitude of their observing station. Under these conditions, summer temperatures become about 3,000° K.

At first glance it would seem that since higher summer than winter temperatures are reasonable the above method provides a simple technique for the evaluation of an ionospheric climate. However, such is not the case, as was first revealed by Berkner, Wells and Seaton (1936) and Seaton and Berkner (1939). These workers showed that during December the F2 layer electron density in *both* hemispheres was greater than that found in June. These observations completely destroyed the Appleton–Naismith hypothesis.

8.5. *Latitudinal* F2 *Region Temperature Variations*

From (8.14) it is also possible to deduce a theoretical latitudinal temperature variation. If the subscripts c and d refer to conditions of the variable at co-latitude θ_c and θ_d respectively, it is possible to obtain

$$n_c/n_d = (T_d/T_c)^{(r+1)/2}[\sin(\theta_c+\delta)/\sin(\theta_d+\delta)]^{1/2}. \quad \ldots\ldots(8.18)$$

While it is possible to use the observed values of electron density at various latitudes for a given solar declination, it hardly seems advisable to solve the above equation in view of present uncertainties regarding the value of the exponent r.

In comparing electron densities over Hiraiso (latitude 36° N.) and Asahigawa (latitude 44° N.), Japan, Maeda and Tukada (1938) observed that when $r=1$, the maximum F2 region electron density over the former station was 20% greater than that over the latter during nearly simultaneous experiments in June 1936. From these data, temperatures over Hiraiso compared with those over Asahigawa would be as 1·0 to 2·2, i.e. temperatures at the more northerly location would be more than twice those at the southern site. Such latitudinal temperature gradients do not appear reasonable.

Table 10. Ratio of F2 Ionospheric Temperatures (given station compared with Trinidad, B.W.I.)

Station	Latitude	January		August	
	° N.	$r=1/2$	$r=-3$	$r=1/2$	$r=-3$
Clyde	76·5	—	—	6·20	0·25
Churchill	58·8	0·44	1·85	3·46	0·39
Trinidad	10·6	1·00	1·00	1·00	1·00

A comparison has been made between latitudinal temperature variations in the F2 region during 1947. Employing monthly mean values at 1800 G.M.T., it was found that during January and August respectively the ratio of temperatures at Clyde, Baffin Island (Canada), Churchill, Manitoba (Canada), and Trinidad, B.W.I., are as given in Table 10.

8.6. *Conclusion*

The various results obtained in this section yield but one inference: *the estimation of temperatures, either laboratory or upper atmospheric, by means of the recombination coefficient is at present unsound.* The entire subject requires extensive investigation before the techniques described above may be employed with any reasonable hope of success. The breakdown of the methods previously listed may be attributed to at least the following facts: (*a*) the non-Chapmanian behaviour of the F2 ionospheric region, (*b*) the difficulty of applying a not clearly understood laboratory relationship to the ionosphere and (*c*) the neglect of the effect of the terrestrial magnetic field upon the upper atmospheric recombination processes. Surely the conclusions of Maeda and Tukada (1938) when referring to this method, "It therefore seems that the theory involving the temperature of the F2 region is unnatural", can hardly be refuted.

§9. TEMPERATURE AND THE INFLUX FUNCTION

9.1. *Atmospheric Contractive Effects*

The use of the influx function for the determination of temperatures in the F2 ionospheric layer arises from attempts to explain the measured nocturnal ionic concentrations. Results of conventional theories require that the electron density during darkness in the F2 region decrease at a greater rate than is experimentally observed. Since non-solar effects producing ionization at these altitudes may be considered as negligible, and since electron attachment processes are usually considered as small (cf. Massey 1940, Woolley 1946, Bates and Massey 1948), an atmospheric contractive effect which tends to preserve nocturnal electron densities has been postulated. As the magnitude of the contractive effect is somewhat sensitive to temperature, it may become possible to make reasonable estimates regarding the ratio of day to night temperatures by use of this method.

A qualitative expansion effect has been proposed in the literature. Appleton (1935) and Hulburt (1938) both explained the differential northern hemisphere winter–summer electron densities by means of an expansion hypothesis. Tani, Ito and Sinkawa (1937, 1938) attributed the F2 layer behaviour to thermal expansion of the atmosphere and changes in solar radiation. Harang (1937) estimated that a diurnal night-to-day temperature range of 1/1·57 was necessary to explain the lifting of auroral heights in the polar sunlit atmosphere. Alpert (1947) concluded that considerable expansion of the F2 region occurs during eclipses. Bannon and Wood (1946) maintained that if the thermal expansion hypothesis be correct, midnight temperatures are probably low and relatively constant throughout the year, and that summer noon temperatures are very much higher than those at midnight. Martyn and Pulley (1936) estimated that the F region cools after sunset by an average amount of 30%, the actual amount of cooling varying from night to night. In attempting to explain the diurnal variation in ionization density, Ghosh (1940) considered the possibility of a temperature decrease in the ionospheric layer. Hulburt (1934) suggested that the expansion of the F2 layer was due to heating causing winds which "blow a wave of ionization with them". Ranzi (1939) stated that he observed a decrease in F2 ionic concentration caused by thermal expansion.

9.2. *Diurnal Chronothermal Function*

It will be assumed that a single temperature–altitude relationship, valid throughout the 24-hour period, exists to an altitude of 100 km. Above this level, both a diurnal and a seasonal effect upon the temperature is envisaged.

To a first approximation an exponential temperature decrease with time is assumed during the period from two hours before sunset until sunrise and an exponential increase in temperature from sunrise until two hours before sunset. A function of this type applies fairly well to conditions at the earth's surface during the same time interval and for lack of more accurate data will be considered to hold at the level of the F2 region. Thus from two hours before sunset until sunrise

$$T = T_m e^{-2at} \qquad \ldots\ldots (9.1)$$

where T_m is the temperature two hours before sunset. From sunrise until two hours prior to sunset

$$T = T_s e^{2bt}. \qquad \ldots\ldots (9.2)$$

The same minimum temperature $T_s = 300°$ K. is assumed to exist throughout the year at sunrise. Considering the *isobaric* surface where P_m is constant, the number of particles during darkness is then given by

$$N = (P_m/kT_m)e^{2\alpha t}. \qquad \ldots\ldots(9.3)$$

Another important factor is the altitude to which a given isobaric surface $P_m = P(T_m)$, originally at 300 km., sinks by sunrise. This lowest altitude is shown in Figure 15 where the ordinate represents the maximum temperature T_m of the isobaric surface $P = P(T_m)$, and the abscissa represents the altitude of this surface at sunrise.

9.3. *Development of Equations*

A quantitative expression of the contractive effect may be easily derived (cf. Gerson 1949 b). Define an influx function

$$s(t) = (n/N)(dN/dt) \qquad \ldots\ldots(9.4)$$

which specifies the proportion of ions of one sign which flow into a unit volume under the influence of an atmospheric contraction. The equation describing the rate of disappearance of electrons is then

$$dn/dt = q + s(t) - \alpha n^2 \qquad \ldots\ldots(9.5)$$

if attachment of free electrons to neutral particles is insignificant.

Since $\alpha = \alpha(T)$, it now becomes possible to write

$$\alpha = AT^r = \alpha_0 e^{-2\alpha r t} \qquad \ldots\ldots(9.6)$$

where $\alpha_0 = AT_m{}^r$. Equation (9.6) becomes during darkness when $q = 0$

$$dn/dt = (n/N)(dN/dt) - \alpha_0 n^2 e^{-2\alpha r t}. \qquad \ldots\ldots(9.7)$$

Quadrature of (9.7) under the assumption that the F2 region lies on an isobaric surface yields, when $r = +1/2$,

$$1/n = (\alpha_0/a)e^{-at} + Ce^{-2at} \qquad \ldots\ldots(9.8)$$

where C is a constant of integration.

According to the discussion given in § 8, it seems advisable to choose for the F2 region a value of $r = +1/2$. This value indicates a mechanism whereby electrons are transferred directly from negative to positive ions when the two approach within some critical distance of each other. In the reaction negative ions are considered to act as dynamic catalysts which modify the process rate. The neutral oxygen atom which results when the electron is lost by detachment binds a new electron at its next electron encounter. Such a mechanism removes the necessity for a three-body collision, which is extremely unlikely in the low pressure environment characteristic of the upper ionosphere.

During daylight, it will be postulated that the level of maximum ionization and the rate of maximum production of ions both lie upon an isobaric surface. From (9.5) and (9.4)

$$dn/dt = q(t) + (n/N)(dN/dt) - AT_s{}^{1/2}n^2e^{bt} \qquad \ldots\ldots(9.9)$$

where the curved-earth value of the ion production $q(t)$, defined in equations (2.18) and (2.20), is taken. Quadrature of (9.9) cannot be accomplished analytically. However, an approximate solution may be obtained by replacing the differential form dn/dt by the incremental form $\Delta n/\Delta t$ and considering $\Delta t = 3{,}600$ seconds.

9.4. *Computational Results*

Graphs indicating the actual and computed nocturnal ionization decrease in the F2 ionospheric layer over Washington, D.C., are shown in Figures 16 and 17. Some extreme values of temperature were used in the computations the better to illustrate the effect involved, and not because such values are possible in planetary atmospheres.

On examining the computed curves, it is obvious that with moderate or extreme diurnal temperature variations in the ionosphere, the effect of an influx of ions (caused by an atmospheric contraction) should become apparent during the latter half of the darkness period. During the early part of the darkness

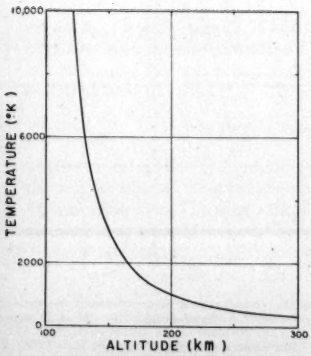


Figure 15. Sunrise altitude of a given isobaric surface $P=P(T_m)$ originally at the temperature T_m and an altitude of 300 km. An exponential temperature decrease during the night to $T=300°$ K. at sunrise is postulated.

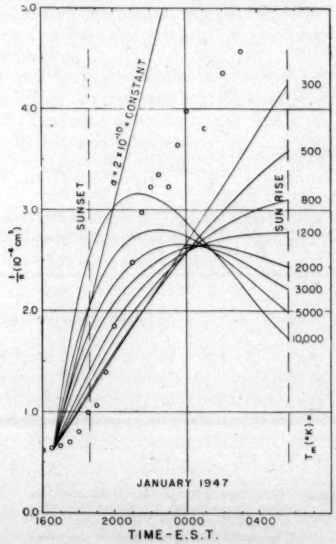


Figure 16. Experimental and computed decrease in nocturnal, F2 region ionization during January 1947 for Washington, D.C. The computed results include the effect of atmospheric contraction from the given temperature at two hours before sunset to the temperature of 300° K. at sunrise.

period, the removal of ions by recombination appears to be the dominating action, but after about five hours the effect of contraction becomes more and more important, depending upon the contractive effect (temperature decrease) considered.

It is seen that the nocturnal experimental data do not fit any of the theoretical curves. The observations appear to indicate that some type of constant recombination mechanism was operative, and that the value of the recombination

coefficient during January was not necessarily equal to that occurring during August. *Indeed, it must be concluded that for the median monthly data chosen, strong diurnal temperature changes were absent in the* F2 *region.*

Gerson (1949 c) had originally found that a nocturnal temperature decrease from 2,000° K. to 300° K. appeared to explain the night-time ionization density during magnetically quiet periods. However, in view of the more general results depicted here, such a temperature variation in general seems unlikely.

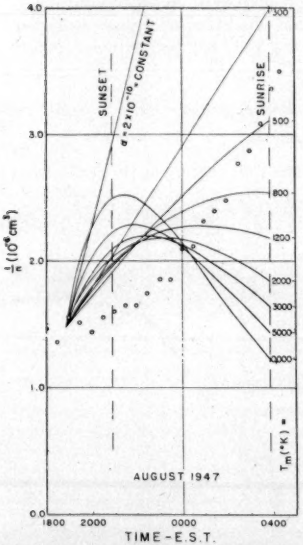


Figure 17. Experimental and computed decrease in nocturnal, F2 region ionization during August 1947 for Washington, D.C. The computed results include the effect of atmospheric contraction from the given temperature at two hours before sunset to the temperature of 300° K. at sunrise.

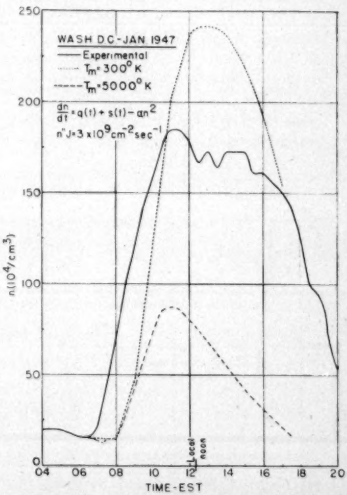


Figure 18. Experimental and computed increase in daylight electron density for the F2 region during January 1947 at Washington, D.C. The result includes in the case of T_m=5,000° K. the effect of electron divergence caused as the atmosphere heats from 300° K. to 5,000° K.

Figure 18 indicates both the experimental and the theoretical results computed from (9.9) for the daylight period during January 1947. For the purposes of calculation it was assumed that in daylight $n''J = 3 \times 10^9/\text{cm}^3$ sec. corresponding to a noon ion production during January of about 450/cm^3 sec. and 115/cm^3 sec. for $T_m = 300°$ K. and $T_m = 5{,}000°$ K. respectively.

In the evaluations, consideration was given to the fact that the height of the isobaric surface, $P(T_m)$ for $T_m = 5{,}000°$ K., changed during the course of the day. The quantity D was obtained by the numerical integration of (2.19).

The theoretical curves indicate the expected result that when an atmospheric expansion is considered, the resulting efflux of ions decreases both the maximum electron density and the rate of change of electron density from that anticipated for an isothermal atmosphere. However, in neither example can it be stated that the evaluated results agree with the observed values. The lag in the increase of electron concentration is a large discrepancy between theory and observation. Further, the rate of increase or decrease of electron concentration does not conform to that experimentally observed. The results for intermediate values of temperature T_m may fit the experimental curves.* Because of the labour of computation, evaluations were not performed for August although it is expected that similar conclusions would be drawn.

The lack of agreement between the calculated and derived values of electron concentration may be attributed again to the fact that a strong diurnal temperature variation probably does not exist in the F2 region.

The above results may be compared with those obtained independently by another investigator under somewhat different assumptions. Hechtel (1941) also considered the effect of an atmospheric expansion upon the rate of change of electron concentration. In place of our (9.9) Hechtel obtained

$$dn/dt = q'_{\max} - \alpha n^2 - (n/T)(dT/dt) \qquad \ldots\ldots(9.10)$$

where $q'_{\max}$ is defined in (2.9) and, of course, $dN/N = -dT/T$. In the solution of (9.10) two cases were considered: (*a*) $\alpha =$ constant and (*b*) $\alpha = CT^{-1/2}$ where C is a constant. For a constant value of recombination coefficient expansion and rearrangement of terms in (9.10) gives

$$dT + \alpha T\, ndt + T\, dn/n = (n''Jmg/nke)(\cos\chi)\, dt \qquad \ldots\ldots(9.11)$$

which was solved for T/T_a where T_a is defined as the value of T at $t = t_a$. Thus the relative diurnal temperature range was evaluated although the magnitude of T_a was unknown. The right side of (9.11) was numerically integrated, and it was assumed that after 24 hours the temperature cycle repeated itself and that

$$\int n dt = \tfrac{1}{2} \sum_{l=0}^{l=24} (n_l + n_{l+1})\Delta t$$

where $\Delta t = 3{,}600$ seconds. Results were given in terms of the ratio T/T_a for various values of the recombination coefficient.

With average F2 ionospheric data for the period 6th–10th July 1939, presumably obtained in Germany, Hechtel obtained the following ratios of maximum to minimum temperature for different values of recombination coefficient α:

α (cm³/sec.)	$T_{\max}/T_{\min}$
$1{\cdot}5 \times 10^{-10}$ or greater	27 or greater
$1{\cdot}0 \times 10^{-10}$	10
$0{\cdot}8 \times 10^{-10}$	6

It was also found that the maximum temperature occurred between

* Any daylight temperature increase must, on the average, balance a nocturnal temperature decrease; otherwise the atmosphere would either continually gain or continually lose heat.

1400–1630 hours local standard time, and that the minimum temperature occurred between 0400 and 0630. For the instance where $\alpha = CT^{-1/2}$, the equation was brought into the form

$$dn/dt = \psi(n, dn/dt, T, \chi). \qquad \ldots\ldots (9.12)$$

Graphical integration by the method of trial and error was used. Results were calculated in terms of the ratio T/T_a, again assuming that the temperature cycle repeated itself after the 24-hour period. Under these conditions, the ratio $T_{max}/T_{min} = 4{\cdot}5$ when $\alpha(T_{max}) = 4{\cdot}3 \times 10^{-11}$. For both conditions of recombination coefficient, the noon ion production is about 60 electrons/cm³ sec.

Even if it is assumed that $T_{min} = 300°$ K. in the F2 region, it is difficult to conceive a ratio where T_{max}/T_{min} is very much greater than 10. In any event the method involving a constant recombination coefficient appears to be extremely critical with respect to the value of α, so that a small change in α results in an apparent large diurnal temperature variation. Thus, errors inherent in the experimental determination of recombination coefficient are grossly magnified in so far as the temperature ratio is concerned.

For the condition where the recombination coefficient is a function of temperature, the ratio $T_{max}/T_{min} = 4{\cdot}5$ appears more reasonable. However, it should be observed that this result is not directly comparable with that given in Figure 18 because of the fact that Hechtel employed (*a*) a flat earth value of q'_{max}, the rate of ion production, and (*b*) a recombination coefficient temperature dependence where $r = -\frac{1}{2}$. In evaluating the results portrayed in Figure 18, the curved earth approximation to the rate of ion production was utilized together with a value of $r = +\frac{1}{2}$.

Several comments may be made regarding the validity of using equations (9.9), (9.11) and (9.12) for the determination of a daylight temperature trend in the F2 region. As this layer is distinctly non-Chapmanian in formation and structure, the ion production formulae for either flat or curved earth should not be applied. (It should also be recalled that the Chapman equation for ion production was obtained for an isothermal atmosphere.) The assumption that the level of maximum ion production lies upon an isobaric surface is open to question and in itself may account for a major discrepancy in the computations. Further, there is insufficient knowledge available to specify the relationship between temperature and the recombination coefficient. Also the recombination coefficient and the ion production usually cannot be uniquely determined from the daily range of electron concentration. Finally, it must be recalled that q'_{max} (in equation (9.10)) is only valid for zenith angles χ less than or equal to 80°.

Because of the reasons given above, it does not appear advisable to use the given equations to determine a possible diurnal temperature change. A much better method undoubtedly lies in the calculation of a nocturnal temperature decrease; in this case the complicating effects involved in the ion production term vanish.

On the basis of equations involving the influx function a moderate or strong daylight temperature change has not been conclusively established. Similarly, a large nocturnal temperature decrease does not appear. Thus it may be concluded that an appreciable diurnal temperature cycle is absent in the F2 ionospheric region.

§ 10. THERMAL EQUILIBRIUM IN THE MESOSPHERE*

Before discussing the results obtained, we must clarify the meaning of temperature in the rarefied atmospheric regions. It has been suggested that the concept of temperature at great altitudes above the earth's surface differs from that employed in the laboratory.

The classical view of temperature is embodied in a law of thermodynamics, viz.: if two assemblies are each in thermal equilibrium with a third body, they are in thermal equilibrium with each other. Thus the condition for thermal equilibrium is the equality of a specific single-valued function of the thermodynamic states of the assemblies. This function is termed the temperature. Any of the assemblages may then be employed as a thermometer to yield a quantity known as the empirical temperature.

When the number of particles in a given assembly is drastically reduced, consideration must be given to fluctuation phenomena and deviations from the average. If Q indicates the standard deviation of energies divided by the average deviation, it may be easily determined (Lindsay and Margenau 1936) that

$$Q=(2/3N)^{1/2}. \qquad \ldots\ldots(10.1)$$

This equation reveals that the spread of energies within assemblies commonly encountered is insignificant; thus under conditions of normal temperature and pressure $Q\simeq 10^{-10}$. At 300 km., Q increases to about 10^{-5} if the temperature at that level is 3,000° K. (see Table 3). While this value of Q is still very small, nevertheless, the probability of deviations from a Maxwellian distribution is very much greater. It should be recalled that with a Maxwellian velocity distribution the various atomic and molecular energy levels are populated and depopulated at the same rate, thus allowing no emission spectra to be observed. The observation of such spectra in the light of the night sky therefore indicates that (assuming the lines are not caused by incoming particles) some deviations from thermal equilibrium do occur (Petrie 1948 b). While the various computational methods should yield the same values in an equilibrium atmosphere, temperatures obtained will vary with the technique employed if thermal equilibrium is not attained. Using observed auroral intensity distributions in a number of nitrogen rotational bands, temperatures were determined to be of the order of 225–240° K. (Vegard 1932, Vegard and Tönsberg 1935, 1938, 1944). The most recent results

* The number density in a planetary atmosphere decreases exponentially (depending upon the temperature–altitude function) with increased radial distance from the surface of the planet. As a consequence, the mean free path also increases until finally an altitude is attained where the mean free path directed radially outward exceeds that directed radially towards the planet. At and above this altitude a gas particle moving outwards will travel farther than a particle moving downwards. Many particles travelling ' up ' will reverse direction only under the influence of the gravitational field, and, before experiencing a collision with another particle, will return towards denser strata of the atmosphere. Thus there exists an outermost or fringe region of the planetary atmosphere where (*a*) collisions between particles are negligible, (*b*) mean free paths are extremely great and (*c*) gas particles move in elliptic, parabolic or hyperbolic orbits depending upon their energies. This region of particle escape is termed the ' exosphere '. The base of the exosphere is known as the ' critical level ' which occurs in the middle of a thick layer from which collisions eject those particles populating the exosphere (see Jeans 1916, Jones 1922, Mitra and Banerjee 1939). The ' mesophere ', or middle atmosphere, extends from the topmost ionospheric layers to the critical level. In the ionosphere and mesosphere of the earth ionized particles in varying percentages are present as a permanent constituent of the atmosphere, and within these regions the aurorae are formed.

The critical level is located at about 1,000 km. above the earth; thus the exosphere extends from about 1,000 km. outwards, the mesophere from about 400 to 1,000 km. and the ionosphere from 70 to 400 km.

were corrected by Nicolet (1943) to values ranging from 250 to 265° K. Harang (1939) obtained 225° K. for regions in darkness and 260–300° K. for regions in sunlight. Vegard and Tönsberg (1941) found temperatures of 220° K. for dark and 240° K. for sunlit aurorae.

The determination of temperature using the vibrational intensity distribution among several emission bands gives abnormally high values. Petrie (1948a) obtained excitation temperatures of 4,600° and 7,600° K. from the relative intensities of helium and hydrogen lines respectively. On examining the intensity of OII spectral lines, Petrie (1947) concluded that the velocity distribution of the exciting particles was characteristic of temperatures from 3,000 to 6,000° K. For OI, 7,000° K. was found (Petrie 1950). Petrie's results also apply to the aurora, which is presumably caused by particle bombardment.

An examination of thermal equilibrium in interstellar space will undoubtedly prove helpful, inasmuch as conditions in such regions represent the extreme to which planetary atmospheres may degenerate. Eddington (1926) indicated that elastic collisions between interstellar particles would establish a Maxwellian velocity distribution with equipartition of kinetic energy among the various atoms. In the interstellar fluid, free electrons may experience with other electrons, ions, or neutral particles elastic or inelastic collisions, the latter causing excitation of the heavy particle concerned. There may also be collisions of the second kind with ions or with atoms or molecules in excited metastable states, free–free transitions in the proximity of nuclei, or capture by the heavy particles. In studying typical planetary nebulae, Bohn and Aller (1947) found that the velocity distribution is very close to Maxwellian, and that deviations from equilibrium were small. Although they concluded that elastic scattering is by far the most probable, it must be recalled that in the assembly they considered the average lifetime of an electron as 10 years—certainly many orders of magnitude greater than is found in the terrestrial mesosphere. A numerical application of their results by Spitzer (1948) showed that deviations from a Maxwellian distribution in interstellar space or somewhat denser nebulae are negligible.

Similar conclusions probably apply to conditions in the tenuous regions of the terrestrial atmosphere where, on the whole, the number of emitting particles is but a small fraction of the total. As this investigation is concerned not with the exosphere but with the earth's ionosphere and mesosphere, it is even more probable that thermal equilibrium exists. It may therefore be concluded with small error that the ionosphere and some portion of the mesosphere is essentially in a state of thermal equilibrium.

§ 11. ELECTRON TEMPERATURES

The discussion would be deficient if a comparison between electron and gas temperatures were omitted. Usually, an energy greater than the ionization energy is absorbed during an ionization process. Because of the disparity in the ionic and electron masses, the excess energy reveals itself as the kinetic energy of the emergent electron. Thus at the instant of ionization, the electron temperature may be defined by

$$T_e = 2B(h\bar{f} - W)/3k \qquad \ldots\ldots (11.1)$$

where $\bar{f}$ is the mean electromagnetic frequency causing ejection of the electron, $B = 7730$ °K./ev., and W is the ionization potential. In collisions with the heavier particles, however, the excess electron energy is rapidly lost even though

insignificant amounts of energy are transferred at each encounter. This fact may be easily determined quantitatively.

Use may be made of an expression derived by Drukarev (1946) relating electron and gas temperatures in the F2 ionospheric region:

$$T_e = T + 3{\cdot}88 \times 10^5(\overline{hf} - W) \qquad \ldots\ldots(11.2)$$

and Cravath's (1930) relationship for the average fractional energy loss per electron collision

$$\Delta E/E = 8mm_e(1 - T/T_e)/3(m_e - m)^2. \qquad \ldots\ldots(11.3)$$

In (11.3) $\Delta E/E$ represents the average energy loss per collision expressed as a fraction of the average electron energy. Since $m \gg m_e$, simplification yields

$$\Delta E/E = +8m_eA/3m(T + A) \qquad \ldots\ldots(11.4)$$

where $A = 3{\cdot}88 \times 10^5(\overline{hf} - W)$.

To indicate the time required after removal of an electron for T_e to become equal to T, the following example may be chosen: $\overline{hf} = 25$ ev., $W = 15$ ev., $M = 20$ and $T = 3{,}000°$ K. Thus $\Delta E/E = 7{\cdot}31 \times 10^{-5}$ per collision. As experimental data reveal that $\nu' = 2 \times 10^3$/sec., $T = T_e$ in about 25 seconds in the F regions and in a smaller time in the E region. Some indication of the high values which T_e might obtain were given by Fuchs (1936a) who found $T_e = 40{,}000°$ K. during periods of high solar activity. Similarly, Barbier (1948) concluded $T_e = 30{,}000°$ K. during an auroral display. However, as is apparent from the example chosen, this energy is quickly distributed among the heavier particles. It may be concluded that the electron temperature is substantially equal to the gas temperature in the ionospheric and lower mesospheric regions.

§ 12. ASSUMED IONOSPHERIC TEMPERATURES

Many values of temperature have been seized upon as representative of ionospheric conditions. Mitra and Rakshit (1938) assumed a linear temperature increase from 300° K. at 100 km. to 1,100° K. at 300 km. Majumdar (1938) assumed $T = 300°$ K. for the E region and 1,000° K. for the F region. Müller (1935) took the E-region temperature in the range 218° to 323° K. and stated "It cannot exceed 100° C. in the E layer". Banerji and Bhatnagar (1938) assumed a more or less linear temperature increase with height from 160° K. at 82 km. to about 1,200° K. at 300 km. Bates and Massey (1946) considered a temperature of 400° K. at 120 km. and 1,000–2,000° K. at the F layer between 200–300 km. Bhar (1938) assumed a temperature of 600° K. above 160 km., and 300° K. between about 100–160 km. Ferraro (1945, 1946), in considering diffusion, postulated $T = 1{,}000°$ K. in the F2 region and 350° K. for the F1 layer. That these and other estimates of temperature in the E ionospheric regions are about the correct order of magnitude is clear from the preceding remarks. However, temperatures chosen for the F layers appear somewhat low.

Das (1936) considered for polar regions a temperature between 100 and 200 km. of 300° K. during darkness and 1,000° K. during daylight. Above 700 km., he indicated that the temperature should be 300° K. throughout the 24-hour period. These values do not agree with the results obtained in this Report.

Some comment should be made regarding the rate of temperature increase in the upper atmosphere. Hulburt (1939) proposed that because of the absorption of ultra-violet radiations by molecular oxygen, the temperature

increased about 50° K/hr. It may be readily determined that if 100 particles/second were ionized in the F region with an average excess of 10 ev., the total excess energy when distributed among 10^{10} particles would create a temperature rise per particle of only about 2·8° K/hr. Such a rate of temperature increase is very low and suggests that other insolation processes or perhaps non-solar agencies may be effective in heating the upper layers.

Regener (1941) attempted to explain high temperatures in the ionospheric regions by qualitatively postulating large ultra-violet absorption coefficients for the major atmospheric constituents. By using these absorption coefficients an appreciable heating of the ionosphere would occur during sunlight, presumably giving rise to some type of diurnal temperature cycle.

Wegener (1943) attempted a determination of ionospheric temperatures on the assumption that the ionosphere behaved as a black body. On this basis he employed the Stefan–Boltzmann law

$$a = \eta T^4 \qquad \ldots\ldots(12.1)$$

and considered $a = 1{\cdot}35 \times 10^6$ erg/sec. = the solar constant. Solution gives $T = 390°$ K. as the maximum possible upper atmospheric temperature. However, if the ionosphere were a black body, the maximum possible temperature must of necessity be lower in order to permit the passage of a considerable portion of solar radiation. As it has already been shown that a temperature greater than this value is necessary to support the observed ion densities, it may be concluded that the upper atmosphere is not a black body and thus does not obey the Stefan–Boltzmann equation.

It seems more probable that the atmosphere follows some type of modified Kirchhoff radiation law. With the usual Kirchhoff law for selective absorption and emission of a gas, it is found that each gas absorbs or emits its own particular frequency. With a mixture of gases, however, the selective heating of any particular component is rapidly transformed by conduction into a general heating of the entire mixture. Thus, if high energy radiation is absorbed by one gas of the atmosphere, all gaseous constituents will be warmed. Selective radiation will first occur, of course, in those gases which emit at the lowest frequencies since their equilibrium at the lower temperature has been disturbed.

§ 13. HEATING BY ELECTRICAL CURRENTS

In addition to a temperature increase arising from the absorption of sunlight in the upper atmosphere, another possible source of heating has been postulated. Close to the turn of the century Schuster (1907) noted that the electric currents in the ionosphere required by his theories were of sufficient strength to produce noticeable heating. These ideas were placed in quantitative form by Chapman (1919) who concluded that heating by current systems postulated by the dynamo theory were unimportant. A re-examination of this problem (Chapman 1937) based on better data revealed that under certain assumptions the rate of ionospheric heating caused by the currents was less than 0·05° K/day. Under more liberal assumptions the maximum rate of heating in the auroral zone was deemed to be about 18° K/hour.

The temperature rise in the atmosphere following a magnetic storm was computed by Peterson (1927) considering given electron currents which penetrated the terrestrial atmosphere. With the assumptions chosen, a considerable temperature increase in the upper atmosphere was deemed possible.

Kirby, Gilliland, Judson and Smith (1935) and later Kirby, Smith and Gilliland (1938) also suggested that magnetic storms may coincide with an abnormal heating of the atmosphere, thus implying an atmospheric expansion on such occasions. Kellogg (1950) also believes that appreciable heating in the upper atmosphere may be attributable to the flow of electric currents. Appleton and Ingram (1935) suggested that the F2 layer may expand because of heating during magnetic storms. Mitra (1947) notes that due to bombardment of the upper atmosphere by fast charged particles there is a further increase of temperature during magnetic storms.

He continues, "During auroral displays, therefore, which generally accompany magnetic disturbances, there is a large enhancement of temperature tending to make the distribution of density with height more uniform. The uniform illumination in auroral streamers may thus be a consequence of uniform density distribution due to heating caused by the impinging particle."

A careful investigation of the various works quoted above leaves the *distinct impression* that the subject is far from settled. Further quantitative work in this field is desirable and necessary.

§14. DIRECTION OF NECESSARY FURTHER RESEARCH

The detailed analysis given in the previous pages attempts to review and explore the most reasonable avenues of deducing temperatures from ionospheric parameters. During the course of the investigation it became increasingly apparent that several specific theoretical and experimental treatments are vitally needed to obtain distinct advances in this field. These deficiencies in current knowledge which apply immediately to improved examinations of ionospheric temperatures are:

(*a*) *Comparison of gas and rotational temperatures.* This problem arises from the consistently lower temperatures deduced from auroral band spectra when compared with temperatures evaluated from all other sources. It has been known for some time (Bonhoeffer and Pearson 1931) that the relative intensity of the OH rotational bands did not correspond to a Boltzmann distribution when excited by discharges through water vapour. The problem was advanced by Oldenberg (1934) who interpreted this phenomenon as arising from the simultaneous interaction of H_2O dissociation and OH excitation with enhancement of the latter effect. Quantitative research was performed by Wakeshima (1942) and Horie (1945), the latter finding that the observed intensity distribution may be attributed to two molecular groups at different equilibrium temperatures. On the other hand, Horie (1943) found that while the rotational energies for both CO_2 and N_2 agreed with a Boltzmann intensity distribution (and thus the gas temperature) the rotational bands of N_2^+ ions indicated temperatures *greater* than the gas temperature. Certainly extensions of the research initiated above, together with re-determinations of the band structures emitted by the aurora are necessary. (There are many well-known experimental examples of band spectra showing higher rotational than gas temperatures. However, spectra showing lower rotational than gas temperatures are observed rather rarely in the laboratory because of the special conditions required for their excitation (Schüler and Haber 1939).) Results of these studies would provide valuable information regarding the excitation conditions of the upper atmosphere.

(*b*) *Determination of cross section for collision of electrons with atomic oxygen.* Additional theoretical studies supported by laboratory work are required to remove ambiguities in the collisional cross section of atomic oxygen with electrons in the energy range 0–0·5 ev. Such research is particularly required because of the apparently great influence of electron–atomic oxygen collisions in the upper atmosphere. The cross section should also be determined for various states of metastability.

(*c*) *Determination of cross section for collision of electrons with various ions.* Both theoretical and experimental work are required to determine the collisional cross section with electrons of the following ions:

$$O^+, O^{++}, O_2^+, O_2^{++}, N^+, N^{++}, N_2^+, N_2^{++}, NO^+, NO_2^+, Na^+.$$

(*d*) *Determination of low pressure recombination coefficient.* Long overdue theoretical and experimental research is necessary in the field of very low pressure volume recombination and its dependence upon temperature. Such studies should include various possible reactions at relatively low energies (0–0·5 ev.).

(*e*) *Investigation of ionospheric sheet currents.* Further theoretical research is needed on the magnitude and characteristics of lateral electron motions in each of the several ionospheric layers.

(*f*) *Diurnal variation of collisional frequencies.* Careful determination of the diurnal trend of collisional frequencies at the levels of the several ionospheric layers at various latitudes is desired as a means for obtaining information on (*a*) diurnal temperature variations and (*b*) latitudinal temperature field at the ionospheric layers involved.

The above list of necessary investigations is not intended to be exhaustive, but if these studies could be concluded, a knowledge of temperatures in the altitude range 100–400 km. would be materially advanced.

ACKNOWLEDGMENTS

The author is pleased to acknowledge the invaluable comments of Drs. R. Penndorf and W. Pfister of the Geophysical Research Directorate, Air Force Cambridge Research Laboratories, and the computational assistance rendered by Messrs. R. J. Donaldson and K. Shure of the same institution. Appreciation is also expressed to Mrs. N. C. Gerson for her valuable assistance during the preparation of this Report.

REFERENCES

Alpert, Y. L., 1947, *C. R. Acad. Sci., U.R.S.S.*, **58**, 1919.
Angenheister, G., 1932, *Terr. Magn. Atmos. Elect.*, **37**, 431.
Appleton, E. V., 1930, *Proc. Roy. Soc.* A, **42**, 321; 1935, *Nature, Lond.*, **136**, 52; 1937, *Proc. Roy. Soc.* A, **162**, 451; 1939, *Quart. J. R. Met. Soc.*, **65**, 324.
Appleton, E. V., and Beynon, W. J. G., 1947, *Proc. Phys. Soc.*, **59**, 58.
Appleton, E. V., and Ingram, L. J., 1935, *Nature, Lond.*, **136**, 548.
Appleton, E. V., and Naismith, R., 1935, *Proc. Roy. Soc.* A, **150**, 685.
Appleton, E. V., and Sayers, J., 1938, *U.R.S.I. Proc.*, **78**, 272.
Bailey, V. A., and Martyn, D. F., 1935, *Wireless Engr.*, **12**, 122.
Banerji, A. C., and Bhatnagar, P. L., 1938, *Indian J. Phys.*, **12**, 387.
Bannon, J., and Wood, F. W., 1946. *Terr. Magn. Atmos. Elect.*, **51**, 89.
Baranul'ko, U. A., 1948. *Priroda*, 34, No. 5.
Barbier, D., 1947, *C. R. Acad. Sci., Paris*, **224**, 635; 1948, *Ann. Astrophys.*, **2**, 193.
Bates, D. R., 1949, *Mon. Not. Roy. Astr. Soc.*, **109**, 215.

BATES, D. R., BUCKINGHAM, R. A., MASSEY, H. S. W., and UNWIN, J. J., 1939, *Proc. Roy. Soc.* A, **170**, 322.

BATES, D. R., and MASSEY, H. S. W., 1943, *Phil. Trans. Roy. Soc.* A, **239**, 269; 1946, *Proc. Roy. Soc.* A, **187**, 261; 1947, *Ibid.*, **192**, 1; 1948, *The Emission Spectra of the Night Sky and Aurorae* (London: Physical Society), p. 48.

BERKNER, L. V., WELLS, H. W., and SEATON, S. L., 1936, *Terr. Magn. Atmos. Elect.*, **41**, 173.

BERNARD, R., 1939, *Phys. Rev.*, **55**, 511; 1941, *Ann. Astrophys.*, **4**, 13.

BHAR, J. N., 1938, *Indian J. Phys.*, **12**, 263.

BIONDI, M. A. and BROWN, S. C., 1949, *Phys. Rev.* **76**, 1697.

BOHN, D. and ALLER, L. L., 1947, *Astrophys. J.*, **105**, 131.

BONHOEFFER, K. E., and PEARSON, T. G., 1931, *Z. Phys. Chem.*, **14**, 3.

BOOKER, H. G., and SEATON, S. L., 1940, *Phys. Rev.*, **57**, 87.

BRANSCOMB, L. M., 1950. *Bull. Amer. Phys. Soc.*, **25**, No. 1, p 27.

BUDDEN, K. G., RATCLIFFE, J. A., and WILKES, N. V., 1939, *Proc. Roy. Soc.* A, **171**, 188.

CHAPMAN, S., 1919, *Phil. Trans. Roy. Soc.* A, **218**, 1; 1931, *Proc. Phys. Soc.*, **43**, 26; *Ibid.*, **43**, 483; *Proc. Roy. Soc.* A, **132**, 353; 1937, *Terr. Magn. Atmos. Elect.*, **42**, 355; 1939, *Proc. Phys. Soc.*, **51**, 93.

CHAPMAN, S., and BARTELS, J., 1940, *Geomagnetism*, Vol. I (Oxford: University Press), p. 529.

CHAPMAN, S., and MILNE, E. A., 1920, *Quart. J. R. Met. Soc.*, **46**, 557.

COWLING, T. G., 1945, *Proc. Roy. Soc.* A, **183**, 453.

CRAVATH, A. M., 1930, *Phys. Rev.*, **36**, 248.

DARWIN, C. G., 1943, *Proc. Roy. Soc.* A, **182**, 152.

DAS, A. K., 1936, *Gerlands Beitr. Geophys.*, **47**, 136.

DIEHL, W. S., 1925, *Standard Atmosphere—Tables and Data*, Report No. 218, (Washington, D.C.: Nat. Advisory Comm. Aer.)

DRUKAREV, G., 1946, *J. Phys. U.S.S.R.*, **10**, 81.

DUFAY, J., 1943, *Ann. Astrophys.*, **6**, 81.

DUFFENDACK, O. S., and CHAO, K. T., 1939, *Phys. Rev.*, **56**, 179.

DUFFENDACK, O. S., REVANS, R. W., and ROY, A. S., 1934, *Phys. Rev.*, **45**, 807.

ECKERSLEY, T. L., 1935, *Nature, Lond.*, **135**, 435.

EDDINGTON, A. S., 1926, *Proc. Roy. Soc.* A, **111**, 424.

ERIKSON, H. A., 1909, *Phil. Mag.*, **18**, 328; 1912, *Ibid.*, **23**, 747.

FARMER, F. T., and RATCLIFFE, J. A., 1935a, *Nature, Lond.*, **135**, 585; 1935b, *Proc. Roy. Soc.* A, **151**, 370.

FERRARO, V. C. A., 1945, *Terr. Magn. Atmos. Elect.*, **50**, 205; 1946, *Ibid.*, **51**, 427.

FISK, J. B., 1936, *Phys. Rev.*, **49**, 167.

FÖSTERLING, K., and LASSEN, H., 1931, *Z. Tech. Phys.*, **12**, 453, 502; 1933, *Ann. Phys.*, **18**, 26; *Hochfrequenztech. u. Elecktroakust.*, **42**, 148.

FUCHS, J., 1936a, *Gerlands Beitr. Geophys.*, **47**, 1; 1936b. *Met. Z.*, **53**, 41.

GEORGE, E. F., 1947, *Proc. Inst. Radio Engrs., N.Y.*, **35**, 249.

GERSON, N. C., 1949a, *J. Chem. Phys.*, **17**, 1351; 1949b, *Nature, Lond.*, **163**, 491, *Ibid.*, 760; 1949c, *Rev. Mod. Phys.*, **21**, 606.

GHOSH, S. P., 1940, *Indian J. Phys.*, **14**, 101; 1944, *Proc. Nat. Inst. Sci, India*, **10**, 333.

GINSBURG, V., 1944a, *J. Phys. U.S.S.R.*, **8**, 253; 1944b, *J. Phys. U.S.S.R.*, **8**, 383.

GINSBURG, V. L., and ALPERT, I. L., 1944, *Izvestiia Akad. Nauk S.S.S.R., Seriia Fizicheskaia*, **8**, 42.

GLEDHILL, J. A., and SZENDREI, M. E., 1950, *Proc. Phys. Soc.* B, **63**, 427.

GONNERMANN, H., and STRANZ, D., 1944, *Aufbau der hohen Atmosphäre bis 400 km.* (Bad Voslau: Zentralstelle für Funkberatg).

GOWAN, E. H., 1928, *Proc. Roy. Soc.* A, **120**, 655; 1930, *Proc. Roy. Soc.* A, **128**, 531.

GRACE, C. H., KELSO, J. M., and MILLER, S. W., 1949, *Basic Ionospheric Research*, No. 3 (Pennsylvania: State College).

GRIMMINGER, G., 1948, *Analysis of Temperature, Pressure and Density of the Atmosphere extending to Extreme Altitudes* (Santa Monica, Calif.: Rand Corp.).

GUHA, U. C., 1949, *J. Geophys. Res.*, **54**, 355.

GUTENBERG, B., 1946, *J. Met.*, **3**, 27.

HARANG, L., 1937, *Terr. Magn. Atmos. Elect.*, **42**, 55; 1939, *Gerlands Beitr. Geophys.*, **37**; 1945, *Geofys. Publ.*, **16**, No. 6; 1946, *Terr. Magn. Atmos. Elect.*, **51**, 381.

HARTREE, D. R., HARTREE, W., and SWIRLES, B., 1939, *Phil. Trans. Roy. Soc.* A, **238**, 229.

HECHTEL, R., 1941, *Hochfrequenztech. u. Electroakust.*, **57**, 108.
HERZBERG, G. L., 1948, *Nature, Lond.*, **161**, 283.
HORIE, T., 1940, *Proc. Phys.-Math. Soc. Japan*, **22**, 787; 1943, *Ibid.*, **25**, 595; 1945, *J. Phys. Soc., Japan*, **1**, 8.
HULBURT, E. O., 1934, *Phys. Rev.*, **46**, 822; 1938, *Ibid.*, **53**, 344, 670; 1939, *Physics of the Earth*, Vol. VIII, Ed. J. A. Fleming (New York: McGraw-Hill Book Co.), p. 492. 1947, *J. Opt. Soc. Amer.*, **37**, 405.
HUXLEY, L. G. H., FOSTER, H. G., and NEWTON, C. C., 1948, *Proc. Phys. Soc.*, **61**, 134.
JACCHIA, L. G., 1948, *Ballistics of the Upper Atmosphere*, Technical Report No. 2, (Cambridge, Massachusetts: Harvard College Observatory).
JAEGER, J. C., , 1947, *Proc. Phys. Soc.*, **59**, 87.
JEANS, J. H., 1916, *Dynamical Theory of Gases* (Cambridge : University Press), Chap. 15.
JONES, J. E., 1922, *Trans. Camb. Phil., Soc.*, **22**, 535.
JONES, M. W. 1950, *Trans. Amer. Geophys. Un.*, **31**, 187.
JONES, M. W., and JONES, J. G., 1950, *J. Met.*, **7**, 14.
KELLOGG, W. W., 1950, *Upper Atmosphere* (Los Angeles, Calif.: Institute of Geophysics, University of California).
KIDSON, E., 1937, *Quart. J. R. Met. Soc.*, **63**, 477.
KIRBY, S. S., GILLILAND, T. R., JUDSON, E. B., and SMITH, N. 1935, *Phys. Rev.*, **48**, 849.
KIRBY, S. S., SMITH, N., and GILLILAND, T. R., 1938, *Phys. Rev.*, **54**, 234.
KNIGHT, J., 1936, *Proc. Roy. Phil. Soc. Glasgow*, **62**, 30.
LINDEMANN, F. A., and DOBSON, G. M. B., 1923, *Proc. Roy. Soc.* A, **102**, 411.
LINDSAY, R. B., and MARGENAU, H., 1936, *Foundations of Physics* (New York: J. Wiley & Sons), p. 234.
LOEB, L. B., 1934, *Kinetic Theory of Gases* (New York: McGraw-Hill Book Co.); 1939, *Fundamental Processes of Electrical Discharges in Gases* (New York : J. Wiley & Sons), p. 131.
MAEDA, K., and TUKADA, T., 1938, *Rep. Radio Res. Japan*, **8**, 21.
MAJUMDAR, R. C., 1937, *Z. Phys.* **107**, 599; 1938, *Indian J. Phys.*, **12**, 75.
MANNING, L., 1947, *Proc. Inst. Radio Engrs., N.Y.*, **35**, 1203.
MARIS, H. H., 1928, *Terr. Magn. Atmos. Elect.*, **33**, 293; 1929, *Ibid.*, **34**, 45.
MARTYN, D. F., and MUNRO, G. H., 1939, *Terr. Magn. Atmos. Elect.*, **44**, 1
MARTYN, D. F., and PULLEY, O. O., 1936, *Proc. Roy. Soc.* A, **154**, 455, and *Radio Res. Board, Australia*, Bull. 101.
MASSEY, H. S. W., 1937, *Proc. Roy. Soc.* A, **163**, 542; 1940, *Bull. Int. Un. Geod. Geophys.*, No. 11, p. 521.
MILNE, E. A., 1924, *Phil. Mag.*, **47**, 209.
MITRA, S. K., 1947, *The Upper Atmosphere* (Bengal: Roy. Asiatic Soc.), p. 407.
MITRA, S. K., and BANERJEE, S. S., 1939, *Indian J. Phys.*, **13**, 107.
MITRA, S. K., and RAKSHIT, H., 1938, *Indian J. Phys.*, **12**, 47.
MITRA, S. K., RAY, B. B., and GHOSH, S. P., 1940, *Nature, Lond.*, **145**, 1017.
MOTT, N. F., and MASSEY, H. S. W., 1949, *Theory of Atomic Collisions* (Oxford : University Press), p. 220.
MÜLLER, A. W., 1935, *Nature, Lond.*, **135**, 187.
NAZAREK, A., 1950, *Bull. Amer. Met. Soc.*, **31**, 44.
NICOLET, M., 1943, *Bull. Acad. Belge Cl. Sci.*; 1948, *Relations entre les phenomènes solaires et géophysiques*, colloques internationaux, Sept. 1947 (Paris : Edit. *Rev. Opt.* (*théor. instrum.*)).
OLDENBERG, O., 1934, *Phys. Rev.*, **46**, 210.
ORNSTEIN, L. S., and HARINGHUIZEN, P. J., 1932, *Z. Phys.*, **77**, 788.
PANDE, A., 1947, *Terr. Magn. Atmos. Elect.*, **52**, 375.
PEKERIS, C. L., 1940, *Terr. Magn. Atmos. Elect.*, **45**, 205.
PENNDORF, R., 1940, *Naturwissenschaften*, **28**, 751; 1946, *Bull. Amer. Met. Soc.*, **27**, 331; 1949, *J. Geophys. Res.*, **54**, 7.
PENNDORF, R., and FLOHN, H., 1942, *Met. Z.*, **59**, 1.
PETERSEN, H., 1927, *Commun. Magnetiques, Copenhagen*, No. 31.
PETRIE, W., 1947, *Canad. J. Res.*, A, **25**, 293; 1948 a, *Ibid.*, A, **26**, 359; 1948 b, *Terr. Magn. Atmos. Elect.*, **53**, 81; 1950, *J. Geophys. Res.*, **55**, 143.
PHILLIPS, P., 1910, *Proc. Roy. Soc.* A, **83**, 246.

PIERCE, J. A., 1947, *Phys. Rev.*, **71**, 698.
RAKSHIT, H., 1947, *Indian J. Phys.*, **21**, 57.
RANZI, I., 1939, *Ric. Sci.*, **10**, 32.
RATCLIFFE, J. A., and SHAW, I. J., 1948, *Proc. Roy. Soc.* A, **193**, 311.
REGENER, E., 1941, *Naturwissenshaften*, **29**, 479.
REGULA, W., 1949, *Met. Rdsch.*, **2**, 267.
ROBERTS, H. E., 1949, *Aero. Engr. Rev.*, **8**, No. 10, 18.
RYDBECK, O. E. H., 1942, *Trans. Chalmers Univ. Tech., Gothenburg*, No. 3, 62; 1944, *Ibid.*, No. 34, 149.
SAYERS, J., 1943, *Rep. Prog. Phys.*, **9** (London : Physical Society), p. 52.
SCHILLING, G. F., 1950, *Methods and Results of Upper Atmospheric Research* (Los Angeles : University of California).
SCHÜLER, H., and GOLLNOW, H., 1938, *Z. Phys.*, **108**, 714; *Ibid.*, **109**, 432; 1939, *Ibid.*, **111**, 484.
SCHÜLER, H., and HABER, H., 1939, *Z. Phys.*, **112**, 614.
SCHUSTER, A., 1907, *Phil. Trans.* A, **208**, 163.
SEATON, S. L., 1947 a, *J. Met.*, **4**, 197; 1947 b, *Phys. Rev.*, **71**, 557; 1947 c, *Ibid.*, **72**, 712; 1948, *J. Met.*, **5**, 204.
SEATON, S. L., and BERKNER, L. V., 1939, *Terr. Magn. Atmos. Elect.*, **44**, 313.
SPITZER, L., 1948, *Astrophys J.*, **107**, 6; 1949, *The Atmospheres of the Earth and Planets* (ed. by G. P. Kuiper) (Chicago : University of Chicago Press), p. 213.
TANI, K. Y., ITO, Y., and SINKAWA, H., 1937, *Rep. Radio Res., Japan*, **7**, 91; 1938, *Proc. Inst. Radio Engrs., N.Y.*, **26**, 1340.
TA-YOU-WU, 1944, *Phys. Rev.*, **66**, 65.
THOMSON, J. J., 1924, *Phil. Mag.*, **47**, 337.
THOMSON, J. J., and THOMSON, G. P., 1928, *Conduction of Electricity through Gases*, Vol. I (Cambridge : University Press), p. 40.
THOMSON, N., 1934, *Proc. Phys. Soc.*, **46**, 436; 1935, *Ibid.*, **47**, 413.
TOSHNIWAL, G. R., PANT, B. D., and BAJPAI, R. R., 1936, *Nature, Lond.*, **138**, 37.
TUKADA, T., 1937, *Rep. Radio Res., Japan*, **7**, 121.
VASSY, A., and VASSY, E., 1942, *J. Phys. Radium*, **3**, 8.
VEGARD, L., 1932, *Terr. Magn. Atmos. Elect.*, **37**, 389, *Geofys. Publ.*, **9**, No. 11; 1939, *Terrestrial Magnetism and Electricity*, Vol. VIII (ed. J. A. Fleming), Physics of the Earth Series (New York : McGraw-Hill Book Co.), Ch. XI; 1946, *Physica*, **12**, 606.
VEGARD, L., and KVIFTE, G., 1945, *Geofys. Publ.*, **16**, No. 7.
VEGARD, L., and TÖNSBERG, E., 1935, *Geofys. Publ.*, **11**, No. 2; 1938, *Ibid.*, **12**, No. 3; 1941, *Ibid.*, **13**, No. 5; 1944, *Ibid.*, **16**, No. 2.
VILBIG, H. F., 1944, *Lehrbuch der Hochfrequenztechnik*, I (Leipzig : Akad-Verlagsgesellschaft), p. 338.
WAKESHIMA, H., 1942, *Proc. Phys.-Math. Soc. Japan*, **24**, 367.
WARFIELD, C. N., 1947, *Tentative Tables for the Properties of the Upper Atmosphere*, TN 1200 (Washington, D.C. : National Advisory Committee on Aeronautics).
WEEKES, K., and WILKES, M. V., 1947, *Proc. Roy. Soc.* A, **192**, 80.
WEGENER, A., 1911, *Thermodynamik der Atmosphäre* (Leipzig : J. A. Barth).
WEGENER, K., 1943, *Gerlands Beitr. Geophys.*, **59**, 278.
WELLS, H. W., and SHAPLEY, A. H., 1946, *Terr. Magn. Atmos. Elect.*, **51**, 401.
WHIPPLE, F. L., 1943, *Rev. Mod. Phys.*, **15**, 246.
WOODWARD, R. H., 1948, *Terr. Magn. Atmos. Elect.*, **53**, 1.
WOOLLEY, R. v. d. R., 1946, *Proc. Roy. Soc.* A, **187**, 403.
WULF, O. R., and DEMING, L. S., 1938, *Terr. Magn. Atmos. Elect.*, **43**, 283.
YAMANOUCHI, T., 1947, *Progr. Theor. Phys.*, **2**, 33.
YAMANOUCHI, T., and KOTANI, M., 1940, *Proc. Phys.-Math. Soc. Japan*, **22**, 61.

SOME APPLICATIONS OF NUCLEAR PHYSICS IN MEDICINE

By W. V. MAYNEORD
Department of Physics, Royal Cancer Hospital, London

CONTENTS

	PAGE
§ 1. Introduction	367
§ 2. The use of the chain-reacting pile for medical purposes	368
§ 3. Fission products	372
§ 4. Comparison of radioactive isotopes from cyclotron and pile	374
§ 5. Medical applications of isotopes	376
(i) Phosphorus	376
(ii) Iodine	381
(iii) Iron	385
(iv) Sodium	386
§ 6. Therapeutic sources of gamma-rays	392
§ 7. External beta-ray sources for therapeutic use	394
§ 8. Dosimetry of artificial radioactive substances	395
(i) Energy absorption and the röntgen	396
(ii) Dosage rates due to radioactive suorces	397
(iii) Dosimetry of beta-ray emitters	398
(iv) Dosimetry of gamma-ray emitters	399
(v) Dosimetry of positron emitters	400
(vi) Fast and slow neutrons	401
§ 9. Medical use of high energy quanta and particles	401
§ 10. Therapeutic use of high speed neutrons	402
§ 11. High energy x-rays	402
§ 12. Protection	407
Acknowledgment	410
References	410

ABSTRACT. A general survey has been attempted of a wide range of applications of nuclear physics to medicine.

A brief description is given of the manufacture of radioactive isotopes in the chain-reacting pile, pointing out some of the limitations of medical importance particularly in respect of specific activity and purity of materials available. A more detailed discussion follows of the therapeutic and diagnostic applications of phosphorus, iodine, iron and sodium. An account is included of the use of cobalt 60 as a substitute for radium in interstitial, intracavitary and telecurie therapy.

The dosimetry of beta- and gamma-ray emitters is discussed in terms of absolute energy units, which alone enable correlation to be made with the previously developed dosimetry of x-ray beams. An indication is given of generalized methods of calculating the energy absorption in tissues and how estimates may be made of the maximum permissible amounts of radioactive isotopes.

The paper includes a review of some of the medical applications of machines generating high energy particles, and in particular describes work by the author and his colleagues with a beta-synchrotron in the region of 20 MeV.

The report ends with a section on protection against ionizing radiations. Some of the biological effects of these radiations are described, methods of estimation of maximum permissible doses are discussed, and a plea is made to workers in the field to co-operate in providing the data which are urgently required to enable more reliable estimates of permissible amounts of radiation to be made.

§ 1. INTRODUCTION

It is now well recognized that recent developments in nuclear physics afford many new opportunities for fundamental medical research and may also provide the basis of improved techniques in practical diagnosis and treatment. It is hoped that the following general account of some of the possibilities opening to us may be of interest.

The applications of nuclear physics in medicine lie in two main fields: (*a*) the use of artificial radioactive substances now produced in such profusion in chain-reacting piles, and (*b*) the utilization of beams of high energy particles or quanta from particle accelerators such as cyclotrons, betatrons, synchrotrons, linear accelerators, and other instruments now available. The first field constitutes a large new region of work inasmuch as many of the new radioactive substances may be ingested or injected into the human body whereas the formerly available radioactive substances, such as radium, are known to be extremely dangerous 'poisons'. The new substances are of course also dangerous, but their chemical nature and short half-lives enable the techniques to be safely developed. Quanta of very high energy allow of a considerable extension of the older methods of using radiation in medicine without promising much which is fundamentally new. Applications in therapeutic and diagnostic radiology constitute one obvious and immediate prospect, but the unfolding picture is in reality much more subtle and entrancing. Since we can build unstable atoms whose chemical properties are those of known elements but whose spontaneous breakdown reveals their presence at the brief climax of their existence, we have a way of following the metabolism of materials in the body and of determining their distributions and activities. We may hope to uncover new secrets of structure and function, to study normal and pathological destructive or repair processes, to attack again the problems of localization of disease or such fundamental but elusive entities as 'shock' or 'inflammation'.

The emission of energy by the new radioactive materials presents us with a destructive agent which, properly controlled, may be of immense value. The use of this energy in the techniques of therapeutic radiology has naturally been one of the first subjects to be developed, but we must guard against the idea that nuclear physics is primarily concerned with therapeutic radiology. Already we see a trend rather towards diagnostic and physiological applications. We are offered new methods of studying disease of the peripheral blood circulation, of studying bone destruction and regeneration, of throwing new light on disease of the thyroid in adults and children. We also have new methods of tumour localization in certain sites, and even possibilities of discrimination between benign and malignant tumours. In gynaecology a fresh approach is possible, in certain instances, to both medical and surgical aspects of the subject; in ophthalmology possibilities unfold of new physiological studies of the eye and new techniques in the treatment of diseases, including tumours, of that organ. The pathologist and physiologist are particularly favoured with techniques for the study and comparison of normal and abnormal, and the haematologist with more subtle ways of following the life histories of blood cells, their development and destruction. Knowledge of the distribution of a wide range of elements throughout the body in health and disease must lead to advances in medical art and science. Already the Geiger or scintillation counter shyly takes its place beside the stethoscope and clinical thermometer. The surgeon,

profiting by advances in physiology and other sciences, may receive vital aid in his difficult decision as to the extent of radical or conservative surgery, and has available new radioactive means of modifying and improving auxiliary radiotherapeutic techniques.

It is obvious, too, that the biochemist, profiting by the new situation, may learn much of the metabolism and ultimate fate of many materials. Indeed, resulting advances in biochemistry and biophysics will probably prove of far greater importance to medicine than the rather obvious direct applications mentioned above. Since the present need of radiation therapy is fundamental understanding rather than new technical gadgets, the use of radioactive isotopes and the intense streams of protons, deuterons, neutrons, positrons and gamma-rays in the study of the fundamental mechanisms of the action of radiations in living tissues may prove of great importance.

There are, too, increased social responsibilities. Radiobiology, confronted with large-scale social problems in both peace and war, must utilize to the full its resources if it is to respond adequately to the heavy new demands. The developments in nuclear physics have indeed provided new and serious problems. Quite apart from any threat of war, in the development of atomic industry groups of workers will be liable to exposure to ionizing radiation of a very dangerous character. During the development of nuclear physics techniques in the atomic energy projects, elaborate protective devices and codes of rules of safety have been worked out, and these rules must now be developed and applied to ensure the adequate protection of the new and often less experienced workers who will serve the industrial, medical and social developments of nuclear applications. Problems of public health in future will include the provision of adequate safeguards against the contamination of water supplies, sewage disposal systems, and, indeed, whole cities, by the radioactive substances now apparently manufactured with ease in atomic energy establishments or research laboratories, and subsequently issued to industry, hospitals, and other institutions. We will return to these problems later and here merely remind nuclear physicists that they have special responsibilities to innocent and inexperienced enthusiasts working with them, as well as to the community at large and to themselves.

§ 2. THE USE OF THE CHAIN-REACTING PILE FOR MEDICAL PURPOSES

We will presume the reader to be aware of the elementary physics of the chain-reacting pile and proceed to discuss its significance from a medical pointof view.

The pile may be regarded as an extremely bulky and complex source (*a*) of fast and medium energy neutrons from fission, (*b*) of slow neutrons in the reactor itself, reflectors and 'thermal columns', and (*c*) of very strong beta- and gamma-rays, chiefly from fission products and their daughter elements.

It is clear that any direct therapeutic use of the pile is at the moment out of the question. Its bulk and immobility, the difficulties of defining beams of radiation, and their mixed nature, preclude its therapeutic use, quite apart from the fact that a pile is likely to be situated in a relatively remote area not easily accessible to universities and large centres of population. Also, the mean energy of the neutrons available is too low for them to have the required penetration in tissues for the treatment of any but very superficial lesions. It will be remembered that neutrons rapidly lose energy in traversing hydrogenous media

and are, therefore, in tissues brought to 'thermal' levels of energy in a few centimetres, the precise depth depending of course on initial energy. The fact that beams of neutrons from a pile are likely to be mixed with intense gamma-radiation would also make the results of application very hard to interpret. It would be difficult to obtain any of the distinctive advantages associated with either type of radiation.

Much interesting biological work has been carried out with the pile, as, for example, studies of effects of whole-body irradiation of animals, but we must exclude the discussion of such studies from our present scope. With the development of small 'enriched' piles more direct application may be possible. At the moment medical interest centres chiefly in the use of the radioactive isotopes made in the pile. These isotopes fall into two main groups, (*a*) those manufactured by placing suitable stable isotopes in the fast and slow neutron fluxes within the pile, and (*b*) the fission products incidental to the chain reaction itself.

In the first group the chief induced activities are due to slow neutrons which predominate in most of the 'holes' into which materials may be inserted. The most common reactions are due to neutron-induced activities formed by:

(*a*) simple neutron capture, yielding a radioactive isotope of the parent element,

(*b*) transmutation reactions such as (n, p) or (n, d) reactions which yield isotopes differing chemically from their parents, and

(*c*) (n, α) reactions which are followed by a decay or series of decays, again yielding new chemical elements.

For medical applications it seems that the simple (n, γ) products are much the most interesting but suffer from the important difficulty that the product of the reaction is an isotope of the parent element and therefore no chemical separation is possible. This may cause difficulty by a consequent lowering of obtainable 'specific activity' (in practice specified as millicuries/gm. of material), resulting in bulky sources unsuitable for some important therapeutic applications. Fortunately, the fluxes of neutrons now available for isotope production (10^{11} to 10^{13} neutrons/cm^2 sec.) are sufficient to yield sources of high enough activities for many purposes. While there are relatively few (n, α) or (n, p) reactions, it happens that some of them are of great medical interest. In Table 1 we give a list of a number of important (n, γ), (n, α) and (n, p) reactions, indicating some of the applications to which the material has been put in medicine.

There is one point of importance which must be realized and which may be of medical significance, namely the concentration of radioactive material. The sequence of events during a nuclear reaction is clear. The stable material captures neutrons and builds up radioactive isotopes which commence to decay spontaneously. Eventually an equilibrium is reached when the rates of formation and decay are equal, but this condition takes a time of the order of three to four half-lives to reach even approximately the activity building up exponentially with time. Long-lived isotopes, which are likely to be of most interest to us as external sources of radiation, therefore require prolonged irradiation to obtain the maximum possible amount, and it is often a nice problem how long to irradiate a given material in the pile. If we irradiate for too short a time the amount produced is small, while if too long we have a poor return per neutron and we shall be very unpopular with both those responsible for the pile and our medical colleagues awaiting their radioisotopes.

Table 1

Method of formation	Active isotope	Half-life	Radiation (mev.) β	Radiation (mev.) γ	Medical application of element
(n, γ)	^{24}Na	14·8 hr.	1·4	1·4, 2·8	Circulation studies and therapy
	^{32}P	14·3 days	1·69		Many diagnostic and therapeutic uses
	^{35}S	87·1 days	0·17		Metabolism of proteins
	^{36}Cl	10^6 yr.	0·66		
	^{42}K	12·4 hr.	3·5	?	
	^{45}Ca	8·5 days	K	1·1?	Studies of bone
	^{45}Ca	180 days	0·3		
	^{55}Fe	~4 yr.	K		Metabolism; blood physiology
	^{59}Fe	44 days	0·26 0·46	1·1, 1·3	
	^{60}Co	5·3 yr.	0·3	1·1, 1·3	Therapeutic source
	^{59}Ni	15 yr.	K?		
	^{64}Cu	12·8 hr.	0·58β^+ 0·66β^-		Mineral metabolism
	^{65}Zn	250 days	0·4β^+ K, e$^-$	1·14	Study of colloids and white blood cells
	^{69}Zn	13·8 hr.	I.T. 1·0	0·439	
	^{72}Ga	14·1 hr.	3·4 0·8	0·84, 2·25	
	^{76}As	26·8 hr.	1·1, 1·7 2·7	0·57, 1·25	Metabolism
	^{82}Br	34 hr.	0·465	0·547, 0·787 1·35	Thyroid physiology and therapy
	^{89}Sr	53 days	1·5		Bone physiology
	^{90}Y	65 hr.	2·5		Metabolism; colloids
	^{95}Zr	65 days	1·0, 0·39	0·73	
	^{108,110}Ag	225 days	1·3	0·6, 0·9	
	^{115}Cd	2·8 days	1·11	0·65	
	^{115}Cd	43 days	1·5		
	^{114}In	48 days	I.T., e$^-$ 2·0	0·19	
	^{113}Sn	< 100 days	K, e$^-$	0·085	
	^{122}Sb	2·8 days	0·81,1·64	0·8	
	^{124}Sb	60 days	0·74,2·45	1·72	
	^{127}Te	90 days	I.T., e$^-$ 0·7	0·086	
	Te^{129}	32 days	I.T., e$^-$ 1·8	0·102 0·3, 0·8	
	^{131}Te	30 hr.	I.T. β^-	0·177	
	^{134}Cs	2 yr.	0·75	0·8	
	^{131}Ba	12 days	K, e$^-$	1·2 (?)	Therapeutic studies; metabolism
	^{140}La	40 hr.	1·4, 2·2	1·63	
	^{182}Ta	97 days	0·53	1·22, 1·13	Therapy
	^{185}W	77 days	0·6	0·48, 0·69	
	$^{192, 194}$Ir	19 hr. 70 days	2·2 Present	1·35 0·3, 0·4	Therapy
	^{198}Au	2·7 days	0·8	0·12, 0·44	Therapy—colloids
(n, γ)	^{197}Hg	64 hr.	K, e$^-$	0·075	
	^{203}Hg	25 hr.	K, e$^-$	0·13, 0·16	
	^{205}	51·5 days	0·3	0·28	
(n, p)	^{14}C	6100 yr.	0·145		Metabolism; photosynthesis
(n, p)	^{33}P	14·3 days	1·69		
(n, p)	^{35}S	87·1 days	0·17		
(n, p)	^{45}Ca	180 days	0·3		
(n, α)	^{3}H	~31 yr.	0·015		Water metabolism; photosynthesis

We may illustrate the problem with a simple calculation. Consider a small block of material containing N atoms per cm³ in a flux of n slow neutrons per cm²/sec. The number of atoms made radioactive per second per cm³ is clearly $n\sigma N$, where σ is the capture cross section for the process considered. At equilibrium this number is equal to the number of atoms disintegrating per second, so that $n\sigma N = N_a\lambda$ where N_a is the number of radioactive atoms per cm³ and λ their transformation constant.

The ratio of active to non-active material is therefore $N_a/N = n\sigma/\lambda = 1{\cdot}414\, n\sigma\, T$ where T is the half-life. As σ and T are fixed, we see that the specific activity, or concentration of active material, depends on the neutron flux and is limited by this flux. We may therefore produce a little very active material by placing the substance in the centre of the pile where n is high but there is not much room (and likely to be lively competition for it), or have a good deal of relatively low concentration material by irradiation in the reflectors or other parts of the pile where there is ample space. We see the possible importance of heavy-water piles which make available high neutron fluxes and, therefore, very concentrated materials.

Generally, it is easy to see that the specific activity in curies/gm. at equilibrium is $1{\cdot}62 \times 10^{-11}\, n\sigma/M$, where M is the mass number of the stable isotope.

We may illustrate the problem very well by considering Co60, likely to be the most promising 'substitute' for radium. It emits nearly monochromatic gamma-rays of moderate energy (1·16 and 1·32 MEV.), a little beta-radiation of low maximum energy, and is a suitable material to make into therapeutic sources. Details of its advantages and disadvantages may be found elsewhere. Its half-life is 5·3 years, admirable from the point of view of permanence of therapeutic use, but requiring long irradiation to produce large specific activity since in one year's irradiation in the pile only 12·3% of its final activity is attained.

To take full advantage of the new sources we require fine wire whose activity is of the order of 1 mc/cm. length. Taking a value of $\sigma = 22{\cdot}5 \times 10^{-24}$ cm², a simple calculation predicts a specific activity of 6·1 curies/gm. at equilibrium in a flux of 10^{12} neutrons/sec. cm². After three months, a reasonable irradiation time, the value is only 180 mc/gm., and it is noteworthy that the original American material had an activity listed as 48 mc/gm. With such a concentration the diameter of a wire to give 1 mc/gm., known from experience with radium to be required to enable effective doses to be given in reasonable times, is about a millimetre, a quite thick rod. Again, if Co60 is to replace radium in the so-called telecurie techniques, in which it is used in the manner of an x-ray tube as an external source, we require, if we are to make a substantial advance in technique, quantities of the order of 100 or even 1,000 curies. This activity should be concentrated in as nearly a point source as possible, and certainly in a volume not greater than about 10 cm³ if sharp beams are to be obtained. We see therefore that again in this instance very high specific activity is required if full advantage is to be taken of the new materials. For some sources the high specific activity is not, of course, required, as for example in the use of P32 as a surface beta-ray emitter or Na24 for ingestion. Nevertheless, it is well to realize this physical difficulty, as it can play a large part in determining the medical utility of a given radioisotope.

We may mention that materials are normally irradiated in small aluminium cans inserted into the 'moderator'. Care must be taken in the choice of chemical

compound in which the atom to be activated is bound, so as to avoid so far as possible neutron absorption in associated elements. It is necessary, in view of the very high capture cross section of some impurities, to use very pure materials. Liquids must be sealed in silica since soda glass gives high Na24 activity, and Pyrex is precluded because of its high boron content. Moreover, for medical application, particularly when the materials are to be injected, purity and sterility is of the highest importance and chemical separation usually desirable. For this reason, for therapeutic use, P32 is preferably prepared from sulphur by an (n, p) reaction with subsequent separation, rather than by the obvious (n, γ) reaction on P31. The latter method is, of course, quite suitable for the preparation of beta-ray sources for external application and enables large quantities of relatively impure material to be made. Similarly, I131 may be made by an (n, γ) reaction on tellurium, producing Te131, decaying rapidly to I131, which is then separable.

There remain many problems in producing radioactive materials for medical use. For example, it may be necessary to manufacture a large number of sources to have simultaneously the same activity within fairly close limits. These sources may be required to have accurately defined specific activities per unit length, and again this activity to be accurately constant along their length. Problems of neutron shielding within the pile become important and may lead to undesirable heterogeneity. We leave aside the purely chemical and pharmacological problems, such as the very troublesome heterogeneities liable to be introduced in phosphate solutions on sterilization by heat.

§ 3. FISSION PRODUCTS

The fission fragments pose a different set of problems from the point of view of the medical user. They consist of nuclei with too many neutrons for stability and, as is well known, in a series of beta-disintegrations reach stability usually some four or five elements higher in the periodic table. The half-lives are usually short, but a few have lives of weeks or months. These fission products emit beta- and gamma-rays and, being formed in very large quantities, offer very powerful sources. To give an example of order of magnitude we recall that the best measure of physical size, services and cost of a pile is the rate at which heat is developed. It is well known that a power of 1 watt is the equivalent of approximately 3×10^{10} fissions of uranium per second. Let us consider a fission product like Ba140, produced in great abundance. After the pile has been running for a few days the Ba, with half-life 12·8 days, will be nearly at equilibrium value, and it is easily calculated that the amount of Ba in a pile of moderate size (running at a thousand kilowatts) will be about 50,000 curies. This barium and its associated lanthanum emit gamma-rays of high energy, so that the shielding problem is indeed serious. A relatively small pile can produce all the radioactive materials likely to be required in a large number of medical research organizations, though certain materials, such as I131, may be required in large quantities as experience enables larger and larger doses to be safely administered. For physiological reasons it may be desirable to give one or two large doses of the order of 100–200 mc. each rather than a long succession of small doses which will be utilized relatively inefficiently.

The fission products occur in two groups of masses, in all some forty mass numbers being of importance. The importance of mass number rather than

atomic number is obvious in this connection since in the successive disintegrations only negative beta-rays are emitted, leading to the formation of an isobar of an element of next higher atomic number. The successive numbers of the beta-ray chains therefore consist of particles of the same mass number, but the chain ascending the periodic table in successive disintegrations.

We see, then, that the fission products consist of atoms of neighbouring elements which can theoretically be separated chemically from each other and therefore prepared in a state of high chemical purity. In practice this may not be easy, for the fission products reside in the uranium bars themselves, and the problem is the separation of the products from the large mass of uranium, plutonium and other transuranic elements. This uranium is extremely 'hot' when first removed from the pile and must be treated with the greatest care, so that the chemist is not usually very anxious to carry out the separation processes without good reason. Just to complicate matters, a number of the fission products fall in the rare earth region and therefore the separations are quite difficult, though tremendous progress has been made in this field. The effort to separate fission products, or at least some of them, is, however, well worth while since the activities available are often very great. It will also be realized that in this instance every atom formed is radioactive, so that very high specific activity is possible. For example, in the ideal a mass number M with disintegration constant λ (half-life T) has possible activity of $1{\cdot}13 \times 10^{13}/MT$ curies/gm. Thus, I131 (with a half-life 8·0 days) could have an activity of 125 curies/mg. of iodine. In practice this is unlikely to be attained and small quantities of inactive material must frequently be added for chemical manipulation, but such concentrated sources are very attractive therapeutically.

The bulk of the fission products are probably unsuitable for medical use chiefly owing to their short lives and predominantly beta-radiation, but many suggestions could be made. For instance, the mixed fission products in a section of uranium rod might be used as a source of gamma-rays for treatment. The uranium might be left in the pile for a few months, removed and 'cooled' for a few weeks to be rid of short-lived products, and then used in a telecurie unit. It might be possible to obtain in this way sources of the order of 100 curies in a few cubic centimetres of metal, and these could be returned for chemical extraction after use as a therapeutic source if required. The advantage would be the relatively large quantity of material available easily and cheaply and the absence of costly and dangerous chemical separation. The advantages of very highly concentrated sources are obvious enough. It might be possible to make very small intense sources of iodine, for example, shielded with 'inactive' uranium metal (uranium has one of the highest known linear absorption coefficients for x- and gamma-rays) for insertion into body cavities for the irradiation of tumours in sites only otherwise accessible with difficulty. If the separation of individual elements proved too costly or difficult, certain groups of fission products might be used together such as the rare earth group, alkaline earth group, volatile group including iodine, xenon, bromine, according to the convenience of the chemist. A glance at a table of fission products suggests the use of various individual isotopes as beta-ray therapeutic agents. Sr89, well known for its selective deposition in bone and production of bone tumours in animals, is perhaps on this account a substance to be avoided, but Sr90 and Y90 are both of interest, the former as a beta-ray standard (half-life 25 years) and the latter as a source

of very energetic (2·5 MEV.) beta-rays. The mass chain 95 includes Zr–Cb, Zr having a long half-life (65 days) and, being near the maximum of the fission yield curve, should be available in abundance. Quite apart from I131, Xe133 as a gaseous radioactive product might be considered as a substitute for radon since its half-life is similar (3·5 days), but its beta-rays and gamma-rays (0·085 and 0·030 MEV.) have low energy. Its possible concentration in a small volume and ease of protection make it, however, an attractive source of X-rays for internal diagnostic and therapeutic use. We might, for example, consider the possibilities of a source of Xe in the mouth and the exposure of a film wrapped around outside the cheek. Some interesting new 'views' might be obtained by the use of concentrated sources in body cavities.

§4. COMPARISON OF RADIOACTIVE ISOTOPES FROM CYCLOTRON AND PILE

It was frequently hinted in the early days that the pile made obsolete the use of the cyclotron in the production of artificial radioactive isotopes, but this fear has proved to be misleading and the two instruments are complementary. The pile cannot make the extensive variety of radioisotopes producible with a cyclotron because the latter makes use of a much greater diversity in energy and type of nuclear bombarding particles. The pile is a copious source of medium energy and slow neutrons, the latter giving rise to large yields of isotopes only by (n, f) and (n, γ) reactions. The cyclotron produces many isotopes having a deficiency of neutrons. The pile tends to add neutrons.

The cyclotron can deliver large energies in a small area which the pile cannot do, and many isotopes may be made in the cyclotron which are of great importance in biology and yet cannot be made in the pile. Examples are Be7 (43 days) of interest in toxicology, C11 (20 min.), useful for human experiments, F18 (112 min.) for dental research, Na22 (3 years) in human and plant physiology, V48 (16 days), V49 (600 days), Mn54 (310 days), Co56 (72 days), Ni57 (36 hours), As74 (16 days), Mn52 (6·5 days), Fe59 (47 days) free from Fe55 (4 years) in high concentration, Ag74 (16 days) of interest in industrial medicine, and astatine, that is, element '85' (At211) (7·5 hours), an alpha-emitter concentrated in the thyroid. It will be remembered that the pile normally produces radioactive isotopes of the irradiated element, and that therefore no chemical separation of active material is possible. This may not matter but may prove important, for example if the human or animal body needs only a very small quantity of the material concerned to 'saturate' it. If so, the metabolism of radioactive material in the body depends on its absolute amount. We shall see that even for iodine the metabolism of the radio-element is determined partially by the total amount administered and whether the thyroid is 'saturated'. The same consideration is of much more importance when only minute amounts, as for example of Be or V, are required by the body. In such cases the radioactive isotope should be absolutely free of 'carrier', i.e. of non-active isotope, and this is normally not possible with material produced in the pile. Some carrier may be necessary or advisable to aid chemical manipulations.

Again, in the pile a mixture of isotopes may be produced one of which has a long life and, therefore, is potentially dangerous, for example the mixture of Fe59 and Fe55, the latter having a half-life of 4 years. Pure Fe59, with a half-life of 47 days, is probably preferable for injection or ingestion on grounds of safety

and can be made in the pile by appropriate choice of reaction. The advantage of the pile is its great intensity of neutron flux throughout a large volume and consequent possibility of producing large quantities of active material. The cyclotron is a relatively feeble source of most isotopes, as may be seen by comparing Hamilton's table (Table 2) and the order of amounts available from American and British piles.

Table 2

Radio-element	Type of radiation	Half-life	Energy of radiation in MeV. Particles	γ-rays	Yield (μc/μa.hr.)	Produced by
$_1H^3$	β^-	30·0 yr.	0·013	No γ	0·1	Be-d-$_1H^3$
$_6C^{11}$	β^+	21·0 min.	0·95	....	500·0	B-d-n
$_6C^{14}$	β^-	>1000 yr.	0·150	....	0·00005	N-n-p (Be-d-n)
$_7N^{13}$	β^+, γ	9·93 min.	1·2, 0·93	0·28	1000·0	C-d-n
$_{11}Na^{22}$	β^+, γ	3·0 yr.	0·58	1·3	1·0	Mg-d-α
$_{11}Na^{24}$	β^-, γ	14·8 hr.	1·4	1·46, 2·0, 3·03	10·000	Na-d-p
$_{12}Mg^{27}$	β^-, γ	10·2 min.	1·8	0·9	500·0	Mg-d-p
$_{15}P^{32}$	β^-	14·30 days	1·69	No γ	200·0	P-d-p
$_{16}S^{35}$	β^-	88·0 days	0·107	No γ	0·1	S-d-p
$_{17}Cl^{38}$	β^-, γ	37·0 min.	1·1, 5·0	1·65, 2·15	1000·0	Cl-d-p
$_{19}K^{42}$	β^-, γ	12·4 hr.	3·5	....	50·0	K-d-p
$_{20}Ca^{45}$	β^-, γ	180·0 days	0·2, 0·9	0·7	0·01	Ca-d-p
$_{25}Mn^{54}$	K, γ	310·0 days	....	0·85	1·0	Fe-d-α
$_{26}Fe^{59}$	β^-, γ	47·0 days	0·4, 0·9	1·0	0·03	Fe-d-p
$_{27}Co^{56}$	K, γ, e^-, B^+	270·0 days	0·4 (β^+)	....	1·0	Fe-d-2n
$_{29}Cu^{64}$	β^-, $^+\beta$, K	12·8 hr.	0·58 (β^-) 0·66 (β^+)	No γ	3000·0	Cu-d-p
$_{30}Zn^{65}$	β^+, K, γ, e^-	250·0 days	0·4 (β^+)	0·45, 0·65, 1·0	0·5	Cu-d-2n
$_{33}As^{76}$	β^-,B^+, K, γ	26·8 hr.	1·1, 1·7, 2·7, β^-, 0·7, 2·6, 5^+	1·5, 2·2, 3·2	10·0	As-n-γ (Be-d-n)
$_{33}As^{74}$	β^-, B^+	16·0 days	1·3 (β^-) 1·9 (β^+)	....	2·0	Ge-d-n
$_{35}Br^{82}$	β^-	34·0 hr.	0·7	0·65	10·0	Br-n-γ (Be-d-p)
$_{37}Rb^{86,\ 88}$	β^-, γ	18·0 days	....	....	10·0	Rb-d-p
$_{38}Sr^{89}$	β^-	55·0 days	1·50	No γ	7·0	Sr-d-p
$_{53}I^{131}$	β^-, γ	8·0 days	0·687	0·4	20·0	Te-d-n
$_{85}^{211}$	K, α	7·5 hr.	6·0	....	1·0	Be-α-2n (32 MeV. αs)

HAMILTON, J. G., *Radiology*, 1942, **39**, 541.

Recent developments of the cyclotron have, again, opened new possibilities. The new giant cyclotron in Berkeley produces deuterons at 200 MeV. and alpha-particles of 400 MeV. energy. So numerous and extensive are the transmutations produced by these particles that any bombarded element will produce perhaps dozens of radioactive isotopes extending over a wide range of atomic numbers. Up to forty neutrons, protons and alpha-particles may be 'evaporated' at once. In addition, fission of many lighter elements, such as bismuth, lead and even platinum, has been observed. A rich field of new isotopes may well still be discovered.

It must be remembered that most of the isotopes in greatest demand in medicine and biology, such as C14, P32 and S35, must be produced by irradiation of materials foreign to the pile, and that existing piles were not designed for this purpose. Doubtless peace-time production piles will one day be built. The applications of radioactive isotopes are better exemplified by a study of a few of them than by lists of innumerable possibilities, and we therefore describe below some of the medical uses of isotopes of phosphorus, iodine, iron and sodium, chosen rather arbitrarily to illustrate techniques already developed.

§5. MEDICAL APPLICATIONS OF ISOTOPES

(i) *Phosphorus*

Among radioactive isotopes of medical interest P32 must have pride of place, for a large section of literature on radioactive isotopes in biology and medicine is devoted to it. It emits only negative beta-rays of moderate energy (maximum energy 1·69 MEV., mean energy 0·670 ± 30 MEV.), and has a convenient half-life of 14·3 days, leaving in the body on disintegration a quite negligible amount of stable sulphur. In therapeutic practice excretion is important and shortens the effective half-life. This distinction between physical and biological half-life is continually of importance. The most energetic beta-rays from P32 are able to penetrate some 8 mm. of soft tissue, but in practice few penetrate more than 5 mm. We must not, therefore, expect to detect phosphorus at a greater depth than a few millimetres by observations with a counter on the skin or to produce significant biological effects at more than this distance from the source.

The therapeutic use of P32 depends upon the role of phosphorus in tissue metabolism, particularly its importance in rapidly growing tissue. Knowing that certain tissues, for example bone, contain relatively high concentrations of phosphorus, it becomes natural to ask whether it may be possible to destroy these tissues selectively by irradiation with beta-rays from P32 or selectively destroy neoplasms infiltrating bone. Moreover, since rapidly growing tissue of all kinds requires phosphorus, might it not be possible to destroy such growths with relatively little damage to normal tissues?

Selective absorption. When P32 is administered to an animal it is to some extent selectively absorbed, and this selective absorption is the basis of its use in the treatment of such diseases as leukaemias, polycythemia vera and lymph node diseases. This differential uptake is by no means simple and depends on several factors. For example, the total amount of phosphorus in exchangeable form in a given tissue, the rate of turnover of phosphorus by that tissue and the rate at which new tissue is formed, all contribute to decide the final radioactive content. Bone, for example, has a high inorganic phosphorus content and consequently takes up a large amount of radioactive isotope. Indeed, after a few weeks the bulk (90%) of the labelled phosphorus still present in the body is to be found in bones. The effect of rate of turnover of phosphorus is illustrated by the marked difference in the accumulation of P32 in liver and brain. The two tissues have roughly the same phosphorus content, but liver takes up a large amount of radio-phosphorus in a given time whereas brain takes up extremely little owing to the slow rate of phosphorus metabolism in the brain. The effect of rate at which cellular multiplication and new tissue formation occurs is illustrated by the fact that neoplastic tissue invariably takes up a great deal more phosphorus than similar normal tissue. The factors influencing the instantaneous concentration of P32 at any point in the body are however many, various and very imperfectly understood.

The effects of pathological conditions may be exemplified by the work of Lawrence and his associates (to whom so much of our knowledge of P32 is due), who studied the distribution of P32 in normal and leukaemic mice (Lawrence and Scott 1939, Lawrence *et al.* 1939, 1940, Tuttle *et al.* 1939). A single dose of phosphorus was administered and animals sacrificed after various time intervals, the activity of the tissues being then determined. Figure 1 shows that there was only slightly more P32 in the bone and liver of the leukaemic animals than in the

corresponding organs of the normal animals, but that the lymph nodes and spleens of the leukaemic animals took up two or three times the amounts in the organs of the normal mice, probably due to rapid laying down of new tissue and increased metabolic activity of leukaemic cells. It was later shown by Tuttle and associates that phosphorus is taken up by the blood stream and used in the rapid synthesis of nucleo-proteins at a much more rapid rate by leukaemic cells than by

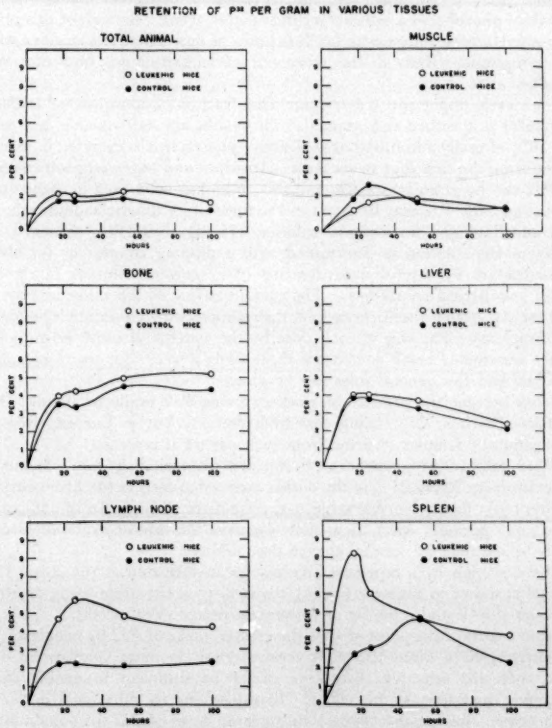


Figure 1. *After* Lawrence and Scott (1939).

cells in the corresponding normal tissues. We cannot here discuss the many investigations into the distribution and uptake of phosphorus in blood cells, but it is worth noting that the distribution of P32 in the tissues of patients who have been treated with radio-phosphorus has been frequently observed, and that leukaemic patients who have received P32 shortly before death usually show great concentration of the substance in bone marrow, lymph nodes, spleen and liver. These findings agree with those in mice. Patients who received P32 many days

before death frequently have the greatest concentration in bone. It appears that the material is first used in the more rapidly metabolizing cells and that later it finds its way into bone. Similar data have also been accumulated from patients with benign and malignant tumours (melanoma, fibrosarcoma, osteogenic sarcoma and cancer of the breast) (Jones *et al.* 1939, 1940a, b, c). The radiophosphorus is frequently preferentially retained in tumour tissues, but it is probable that, as we shall see, the tendency is not great enough to make radioactive phosphorus a satisfactory therapeutic agent. Enough of this material is deposited in normal tissues and cells to limit the dose which can be given without toxic symptoms. Even in the treatment of leukaemia we find that similar difficulties arise.

It is clearly important to determine what fraction of administered radioactive phosphorus is absorbed and retained. The results are very variable, but roughly 15 to 50% of orally administered radioactive phosphorus is excreted in urine and faeces during the first four to six days. Absorption is, however, good enough, so that P32 can be given by mouth, and this route has been used in many studies. For example, the P32 may be given in the form of its dibasic sodium salt, water being added to give an isotonic solution (15 mg. Na_2HPO_4 per cm^3). The activity of the solution is determined with a dipping counter or for absolute standardization by careful measurements of evaporated aliquots in rigorously defined geometrical conditions. The initial activity of the solution may be of the order of 0·20 to 0·4 millicuries/cm^3, the solution being used until it has decayed to 0·05 mc/cm^3. The size of each dose by the parenteral route is some 0·1 to 2·5 mc., treatments being sometimes three times a week, but the dosage is very individual and few general rules can be given.

It has become the practice of workers giving P32 orally to assume 75% of the dose absorbed, 25% being lost by excretion; but in our experience the investigation of samples of urine from each patient is essential.

When radioactive phosphorus is injected intravenously into patients with leukaemia, from 5% to 25% of the dose is excreted, mostly in the urine during the first four to six days. Normal subjects excrete more, of the order of 25% to 50%, in the same period. After these first few days the rate of excretion decreases rapidly to less than 1% of the injected dose per day.

The excretion by a representative patient investigated at the Royal Cancer Hospital is shown in Figure 2 (*a*), (*b*), the activity of the urine being determined by means of a 'liquid' counter of convenient design (Veall 1948).

From a therapeutic point of view the greater intake of P32 by neoplastic tissue is of interest since these cells thus receive relatively more irradiation. It was hoped that this selective absorption might be sufficient to enable effective therapeutic measures to be taken. Investigations in mice and in monkeys demonstrate, however, that with relatively large doses of P32 the bone marrow is destroyed. With sufficient doses, all elements, including the platelets, white cells and red cells, are affected. Thus in the treatment of leukaemia, a disease involving the production of very large numbers of white blood cells, unless white cells are much more sensitive than the other elements in the marrow, it is not possible to prevent their rapid production without at the same time inhibiting red blood cell and platelet formation. It is true that rapidly growing cells such as leukaemic cells take up more P32 than other cells, but if they are are adjacent to other types of cells the latter will also receive irradiation since the beta-rays from

P32 penetrate several millimetres in tissue. Thus even though there is some selective irradiation, especially by virtue of P32 localization in infiltrated lymph nodes, spleen, and liver, one eventually faces the same problem as when using x-irradiation since the whole bone marrow is being irradiated. These effects of P32 on the various elements of the marrow constitute the limiting factor in its use in the therapy of leukaemia and make its use of doubtful value in lymphosarcoma and allied diseases.

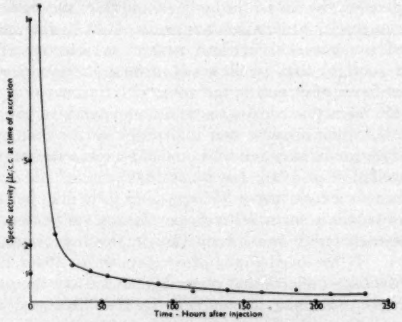


Figure 2(*a*). *After* Mayneord (1950).

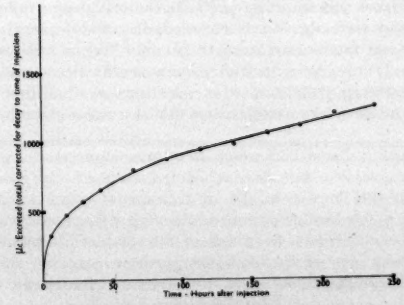


Figure 2(*b*). *After* Mayneord (1950).

It may be possible to cause relatively greater deposition of P32 in other tissues than in the bone marrow and thus decrease the bone-marrow irradiation. In the general problem of the radiation therapy of neoplastic disease one should seek a radio-element or compound of an element which localizes to a high degree in or immediately around the neoplastic cell. There are examples of such localization in normal tissue, as for example in the case of radioactive iodine in the thyroid gland and chromic phosphate (see below) in liver and spleen. In these two instances it has been possible to destroy, in animals, the thyroids and spleens by selective irradiation without serious damage to the rest of the body. These

are examples of selective irradiation in the true sense of the word. Since leukaemia is such a diffuse disease, the possibilities of finding a method of true selective irradiation are not great, and we may have to look for the control of this disease by some method other than by irradiation.

It appears that the degree of selective absorption or concentration required for useful therapeutic techniques with a given element is high. Calculation shows that the factor representing relative concentration in the tissue to be selectively irradiated compared to the rest of the body should be of the order of one hundred, not merely two or three; otherwise therapeutic doses in the desired locus must be accompanied by serious deleterious general or selective effects elsewhere. The concept of 'integral dose' or the study of total energy absorbed in the body helps in a quantitative approach to the subject.

It lies outside the writer's province and competence to assess the results of therapy with P32, but it appears that it is probably the best therapeutic agent available for polycythemia vera and the equal of x-rays in the treatment of chronic leukaemia (Reinhard *et al.* 1946, Lawrence 1949).

Although the therapeutic use of P32 appears a little disappointing, the isotope more than compensates in many other ways. It may yet be that other conditions may respond unexpectedly, and in any case interesting diagnostic possibilities open up for us. It has been found (Low-Beer *et al.* 1946) that a small area of skin with cutaneous nodules may show higher activity than normal skin and that the variations with time of this activity are different. The phosphorus content may even be an index of rate of growth and 'malignancy'. For example, patients with carcinoma of the breast after a dose of the order of 1 mc. of P32 reveal at operation a concentration of P32 in the malignant tumours. The writer and his colleagues have also found that secondary nodules on the skin are about two to three times more active than the normal skin as measured either with a Geiger–Müller counter or scintillation counter (Mayneord and Belcher 1950). Obviously important possibilities of investigating spread of tumours are available to us, and work has been carried out on the treatment of secondary growths by this method.

Another application of P32 which is of great interest arises from studies of the fate of insoluble colloidal particles injected into the body (Jones, Wrobel *et al.* 1944, Allen *et al.* 1945). Some 90% of the material is found in liver and spleen, the latter being highly radiosensitive and showing severe reaction in animals. The distribution is very different from that of the soluble material and there is now evidence that the size of the colloidal particles plays an important role in deciding the distribution in the body. Preliminary experiments have been made on the irradiation of tumours with these insoluble colloids, and also experimental work on animals using colloids of various materials such as Zr, Cb, or Au (Dobson *et al.* 1949, Dougherty and Lawrence 1948). These investigations are merely preliminary, but evidently provide the basis of studies to be made in the future of chemical compounds intended to be retained in tissues to be destroyed.

P32 has also been used in important studies, particularly by Hevesy (1948) and his colleagues, on the metabolism of nucleic acids known to play an important part in cellular reproduction. It has also been used in such problems as the study of the effects of vitamin D in the healing of bone lesions (Morgareidge and Manly 1939, Cohn *et al.* 1939).

The use of active phosphorus incorporated in plastics as a source of beta-rays for the treatment of superficial lesions is described below (see page 394).

(ii) *Iodine*

The study of thyroid physiology and the therapy of thyroid disease has proved one of the triumphs of the use of radioactive substances. The thyroid gland is unique in its ability to accumulate iodine selectively in relatively large quantities, and the concentration in the gland may even reach a value ten thousand times greater than the average throughout the body.

Normal iodine consists of a single isotope (53I127), but the element has a wide range of radioactive isotopes ranging from I122 to I137. Early studies of iodine were made with I128, but its short half-life of 25 minutes prevented its use over times greater than two or three hours and the now readily available I131 is almost universally used in tracer studies or therapeutic applications. Its half-life of 8·0 days is very convenient. It emits both beta- and gamma-rays, the former of maximum energy 0·6 MeV. (mean energy approximately 0·2 MeV.) and the latter chiefly a gamma-ray of 0·367 MeV. plus other radiations, including a hard gamma-ray of approximately 0·6 MeV. The disintegration scheme of I131 has been much discussed, but the latest data are probably sufficiently reliable for clinical use (Cork *et al.* 1949). It is indeed surprising how many disintegration schemes are uncertain, an unfortunate fact from the point of view of medical application, for, as we shall see, the dosimetry of a given substance must remain in doubt until its mode of disintegration is certain.

As we have noted above, I131 is a decay product of Te131 which is produced either by bombardment of Te in a cyclotron with deuterons or by a (n, γ) reaction in the pile.

Hertz, Roberts and Evans (1938) and Hertz *et al.* (1940) first demonstrated the rapidity of iodine accumulation in the thyroid, using rabbits as their experimental animals, and detecting iodine in the thyroid within a few minutes after the administration of the 'tagged' element. It was also noted that the uptake was much larger in the thyroids of animals with hyperthyroidism. These observations were confirmed by Hamilton and Soley in patients with hyperthyroidism. The reader is referred to Hamilton's (1942) excellent account of the use of radioactive tracers in biology and medicine.

An important advance lay in the use of gamma- rather than beta-rays to detect the iodine. In all early experiments the animals had to be sacrificed, ashed or finely minced tissue being then tested for beta-activity. Hamilton and Soley, on the other hand, compared the uptake of labelled iodine by the thyroids of normal controls and of patients with various thyroid disease, *in situ*. This was accomplished by measuring the gamma-rays from the atoms of iodine accumulated in the thyroid. A solution of 24 μc. to 100 μc. of radio-iodine and a total of 14 mg. of iodine in the form of sodium iodide was administered orally to each patient. The radioactivity of the thyroid of each subject was determined by placing a Geiger counter against the neck and directly over the isthmus of the gland, frequent determinations of the labelled iodine content being made over varying periods of time. The radioactivity of the gland could be roughly estimated as a fraction of that of the whole sample administered.

Since these early experiments the techniques of measurement have been considerably refined, and by careful study of 'phantoms' (that is, model thyroids embedded in models of the head and neck) it has been shown how reasonably accurate estimates of amount of material in the thyroid can be made (Mayneord and Belcher 1950). By the use of Geiger counters shielded so as to have strong directional differences of sensitivity, and by careful plotting, the volume of tissue

containing the iodine can be determined. We have also used scintillation counters effectively in such clinical measurements and developed the theory of the measurements.

Normal controls and patients with hyperthyroidism, non-toxic goitre and hypothyroidism were studied by Hamilton, with the results summarized in Figure 3. The ordinate indicates the percentage of uptake of the administered

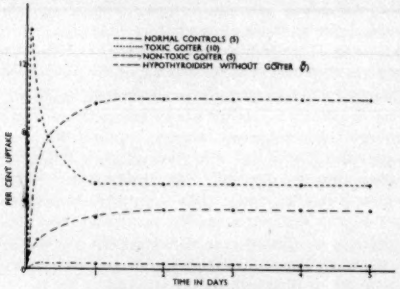


Figure 3. *After* Hamilton (1941).

iodine by the thyroids, and the abscissa represents the time in days after the labelled iodine had been given. The uptake curves illustrate the characteristic pattern of iodine accumulation by the thyroids of the four clinical types included in the series. The relatively large uptake of labelled iodine by the thyroids of patients with non-toxic goitre as compared with the normal controls may be attributed in part to the apparent need for additional iodine by goitres of this type. The form of the uptake curves suggests that the mechanisms of iodine accumulation and storage by the thyroids were similar in these two groups. The prompt accumulation of labelled iodine by the thyroids of patients with hyperthyroidism was followed by a rapid loss of from one-half to four-fifths of the iodine which had been taken up during the first few hours of the experiments. These results were interpreted by Hamilton as indicating not only that the hyperactive thyroid cells had a marked avidity for the accumulation of iodine, but also that the mechanism for the retention of the recently deposited iodine had been altered. The small uptake of iodine by the thyroids of patients with hypothyroidism may be explained by the fact that the thyroid tissue was unable to supply a sufficient amount of the thyroid hormone to meet metabolic requirements; in other words, the failure of these thyroids to accumulate iodine was related to their inability to synthesize the thyroid hormone.

These experiments were repeated with a smaller group of patients and normal controls under the same conditions with the exception that each subject received a test dose which contained a total of approximately 0·1 microgram of iodine. The uptake curves for this group are shown in Figure 4. A marked increase in percentage uptake occurred in all four clinical types. This result was expected since the total iodine content of the thyroid is relatively small and its capacity to accumulate large amounts of iodine is obviously limited. In patients with hyperthyroidism the uptake curves failed to repeat the form they took in the previous study.

The increased uptake of radio-iodine in all groups in which the smaller dose of iodine was used emphasizes the importance of avoiding the dilution of the radio-element with the stable form of the element when it is desirable to prevent the physiological action of flooding the tissues with the administered material. The results of these experiments revealed two effects when the smaller test dose of iodine was employed. First, the percentage uptake was remarkably elevated; secondly, according to the authors the metabolic pattern was entirely different in the group of patients with hyperthyroidism. The larger test dose of iodine was employed in the first series because it was deemed necessary to compare the iodine metabolism of the thyroid in different clinical states under these experimental conditions, since iodine is extensively used in the treatment of both endemic goitre and hyperthyroidism.

A study of the localization of radio-iodine in cancerous thyroid tissue by Hamilton and Soley (1940) and Hamilton, Soley and Eichorn (1940) was initially limited to four cases, in none of which a significant deposition of the labelled iodine in the malignant areas of the thyroid occurred. A comparison of the

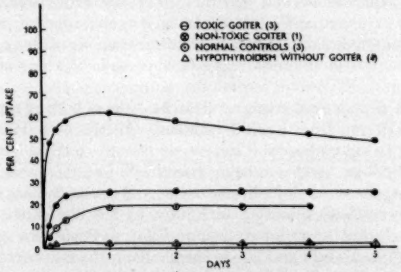


Figure 4. *After* Hamilton (1941).

uptake of radio-iodine in normal and cancerous regions of the gland by radio-autography revealed that the malignant areas took up very little of the labelled iodine.

It would take us too far afield to consider the many contributions to the study of thyroid physiology made possible by the use of radioactive iodine. Biochemical investigations have thrown much light on the metabolism of the iodine in relation to its incorporation into thyroxine and di-iodotyrosine, the two compounds in which it is incorporated in the thyroid gland. Reviews of this subject by Salter (1940) and Elmer (1938), and an excellent account by Leblond (1948) may very profitably be consulted.

Therapeutic use of radioactive iodine. The basis of the therapeutic use of iodine is to be found in studies of animals which have received large doses of iodine. Investigations show that in animals it is possible to deliver to the thyroid doses which produce marked changes in that organ without producing pathological changes in spleen, liver, kidney, pituitary, or any other important tissues. Frequent blood counts reveal little, if any, change due to the radioactivity. Experiments demonstrate in an interesting way the very marked selective irradiation.

The results of the experiments with labelled iodine bring out three points of interest in consideration of the possible use of radio-iodine in the treatment of hyperthyroidism. They show that the thyroids of patients with thyrotoxicosis accumulate and retain a large proportion of the administered radio-iodine; that the thyroid tissue can be destroyed selectively without apparent damage to the other tissues of the body; that accumulated iodine is deposited selectively in the regions of the thyroid tissue which are the most hyperplastic. The areas of hyperplasia, to which presumably the excessive elaboration of the thyroid hormone is mainly referable, would be subjected to the greatest radiation from the beta-rays of accumulated radio-iodine. The selective irradiation of these areas should produce a relatively greater depression of the hyperplastic portion of the gland than of the normally functioning regions. It was deemed inadvisable by Hamilton and his colleagues to employ radio-iodine as a therapeutic agent in hyperthyroidism until the experimental studies on the action of large doses had been completed since, unlike leukaemia and other malignant disorders, hyperthyroidism is essentially curable, and the previous methods of treatment were relatively safe. It was felt that for this reason every precaution should be employed before a potentially lethal agent, such as radio-iodine, was used to treat this relatively benign disease. This admirable caution is greatly to be commended, but the therapeutic use of radio-iodine in hyperthyroidism is proceeding quite rapidly.

In 1946 two papers on the subject described the results of treatment of some fifty patients suffering from hyperthyroidism. In the first, Hertz and Roberts (1946), administering radioactive sodium iodide by mouth, gave doses up to 28 mc., normally 5 mc. to 25 mc. being effective. The therapeutic effect is to be ascribed very largely to the short-lived isotope, and the equivalent dose in röntgens was estimated as equivalent to 500 r. to 2,500 r. ($\pm$ 50%). On the basis of a series of animal and clinical experiments using radioactive isotopes as tracers in the study of thyroid physiology and iodine metabolism, the treatment of twenty-nine cases of hyperthyroidism with initial irradiation by radioactive iodine was instituted. The procedure was carefully studied and it was concluded that ordinary iodine therapy after the administration of radio-iodine offers many advantages in the clinical care of the patients and in the economy and safety of the procedure.

By an analysis, over a long period, of both failures and successes in this series of twenty-nine cases it was shown that radioactive iodine, when given in the dosage range above to uniodized patients with hyperthyroidism possessing goitres of 60 to 75 gm., is highly effective as a cure of the disease in about 80% of cases. When appreciable activity has been administered and subtotal thyroidectomy is resorted to, myxedema or hypometabolism may be expected to develop in a large fraction of the cases (100% of five cases in this series).

The second paper, by Chapman and Evans (1946), describes the treatment of twenty-two cases of hyperthyroidism, no other form of treatment being given. 14 mc. of radio-iodine is calculated as equivalent to 3,490 r. due to the 12-hour isotope. Fourteen patients responded well to a single dose of radioactive iodine, three were given two doses and five were given three doses. Myxedema followed in four patients. Reactions resembling radiation sickness were observed in six patients following large doses, and fibrosis of the thyroid after treatment was also observed in biopsy of two patients. Chapman and Evans do not consider it necessary to administer ordinary iodine with the radio-iodine.

In spite of the small number of cases and the evident disagreement on dosage and treatment detail, it was clear that the use of radio-iodine in thyrotoxicosis was well worthy of further study and a number of recent workers have reported favourably on the use of radio-iodine in the treatment of diseases of the thyroid though, as pointed out by Rhoads (1946), the possibility of tumour production by this method must be kept in mind.

We have seen that cancers of the thyroid do not generally concentrate radio-iodine, but it was found by Keston *et al.* (1942) that radio-iodine may be stored in a metastasis (secondary deposit) from a thyroid carcinoma. Leiter *et al.* (1946) studied two cases of adeno-carcinoma of the thyroid with functioning metastases and hyperthyroidism, the radio-iodine being used as an indicator of the physiological activity of the metastases. In such a case a Geiger counter serves as a method of localization of the secondary deposits. In one patient the absence of thyroid tissue in the neck following surgical removal of the thyroid, confirmed by lack of uptake of radio-iodine in this region, proved that the metastases were the cause of the hyperthyroidism. Thiouracil was found to suppress hormone production in these metastases just as readily as it inhibits the process in the hyperplastic thyroid of ordinary Graves' disease. The patient was treated by the use of large quantities of radio-iodine and a pronounced therapeutic effect was observed. Although the tumour was still present some time later, the disease was under control and the tumour no longer active (Seidlin, Marinelli and Oshry 1946, Marinelli, Foote *et al.* 1947). Reference should also be made to investigations on this subject by Pochin (1950).

It has now been shown that of the several forms of cancer of the thyroid only about 15% pick up radio-iodine selectively, the more malignant the type of tumour the less the ability to store iodine selectively. However, a very interesting beginning has been made and efforts to incorporate radioactive iodine in organic compounds may well result in still more successful techniques.

A study of five cases of adenoma of the thyroid (Leblond *et al.* 1946) established that a considerably smaller amount of radio-iodine was fixed in the adenomatous tissue than in the surrounding normal thyroid. Moreover, thyroid carcinomata show distributions varying a great deal according to the structure of the tumour. Published reports suggest that the less differentiated the tumours, the less iodine fixed. Even if therapeutically ineffective, the use of radio-iodine which is concentrated in the metastases, particularly following surgical removal of the thyroid, enables the spread of the disease to be studied, and sometimes unexpected improvement is observed. In unpublished investigations the writer and his colleagues have observed, for example, encouraging results upon a patient suffering from widespread secondary deposits who has remained well and active for about eighteen months, having been initially quite hopeless by any other technique. There can be no doubt that a very valuable new weapon in the treatment of thyroid disease is to hand.

(iii) *Iron*

Among the elements present in the body iron has long been known to play a particularly important role, yet its metabolism and movement were until recently still largely unknown. It is difficult to summarize the results of researches carried out during the last few years, but an attempt will be made to do so.

Normal iron consists of a mixture of four isotopes, Fe54 (6·04%), Fe56 (91·57%), Fe57 (2·11%) and Fe58 (0·28%). The most used radioactive isotope

of iron is Fe59, emitting gamma-rays of energies 1·10 and 1·30 MEV. with beta-rays 0·26 and 0·46 MEV. maximum energy. Fe59 has a half-life of 47 days and is generally obtained mixed with Fe55 with a half-life of about four years.

The most important questions regarding iron metabolism concern its rate and mechanism of absorption from the alimentary tract (Hahn *et al.* 1938, 1939, 1943, Hahn 1948). It was early discovered by the use of radioactive iron that it was absorbed only in traces by a non-anaemic animal but in abundance by an anaemic one depleted of iron. The mechanism of this absorption is still in dispute but much has been learnt, some of which seems of clinical significance. It has been shown that iron fed to animals or humans appears first in the blood plasma and later in the red blood cells. The mechanism of excretion of iron is also of great interest and it seems that iron once incorporated is excreted only very slowly (Greenberg *et al.* 1943).

Ingenious use has been made of Fe59 to study effects depending on the age of the blood cells. These studies are complicated by the fact that the haemoglobin iron of red cells is derived from the iron of old cells broken down by normal wear and tear, rather than from reserve stores. It is therefore not feasible to use the iron isotope to determine the life cycle of a red blood cell, but if a comparatively small amount of iron is given in a single dose, the age of the red cells into which it has been incorporated can be estimated to a few days. This enables one to distinguish between cells of known ages, varying from a few days to several months, and the general run of cells in circulation. Several physiological investigations involving the study of the variation of fragility of blood cells against various agents with age have thus been made. Interesting new techniques for studying blood volume have also been developed. Ross and Chapin (1943) and Gibson and associates (1947a, b, c) have also made use of these techniques in war studies of surgical shock and the development of techniques for the preservation of whole blood for transfusion. In addition, a number of studies of iron metabolism have been made on patients suffering from pathological conditions such as pernicious anaemia, polycythemia, as well as in normal and abnormal pregnancies. As an example of the kind of information obtained, it has been established that some types of infection seriously affect the absorption of iron by the body, an observation linking up with the clinical experience that anaemias, often refractory to treatment, are associated with infectious and inflammatory diseases (Gibson and Finch 1948).

So far little use has been made clinically of radio-iron, but it requires little imagination to see how physiological studies of the kinds indicated above might in the future affect clinical practice.

(iv) *Sodium*

One of the most interesting sets of observations which promise to be, and indeed already are, of both diagnostic and therapeutic interest, relates to the use of radioactive isotopes of sodium. Sodium is one of the commonest elements in the body, but surprisingly little appears to be known with certainty of its metabolism.

Normal sodium consists of a single isotope, 11Na23, and has two extremely useful radioactive forms, namely Na22 which has a half-life of 3·0 years, emitting positrons of maximum energy 0·58 MEV. and gamma-rays of energy 1·3 MEV., while Na24 is an emitter of both negatrons and gamma-rays, the former of

maximum energy 1·4 MEV. and gamma-rays whose energies are 1·38 and 2·76 MEV. respectively. Two other isotopes are known, Na21 with a half-life of only 23 seconds, and Na25 with half-life 62 seconds, but neither of these two materials would appear to be of any practical importance in medicine. Na22 cannot be made in the pile but is conveniently prepared with a cyclotron.

Na24 is formed by either a (d, p) reaction in the cyclotron or an (n, γ) reaction on Na23 in the pile and in other ways, such as by an (n, p) reaction on Mg. The Na24 transforms by beta-emission, as we have seen, into Mg24, which then emits the gamma-rays. Na22 transforms by positron emission into stable Ne22.

Na24 has a particular importance in that Na23 has a large cross section for neutron capture so that quantities of it are easily made, and sources of the order of many curies can be available if required. Its short life made it appear at one stage a possible substitute for radon.

The clinical use of Na24 commenced in 1936 (Hamilton and Stone 1937), for it was early felt that the disadvantages of internal radium therapy could be avoided by the use of sodium inasmuch as the latter is not fixed in the body tissues and the duration of its effect is limited by the short half-life of 14·8 hours. The initial dose given was large, namely 11 mce. (the mce. 'millicurie

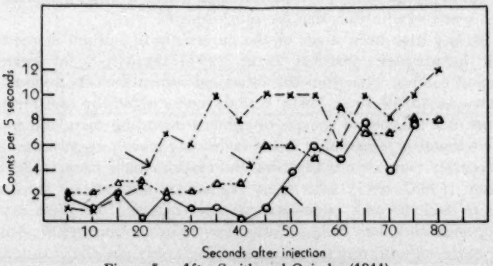


Figure 5. *After* Smith and Quimby (1944).

equivalent' being determined by the comparison with the gamma-rays of radium in a gold-leaf electroscope). It was found by Hamilton and Stone that the decay curve of the sodium could easily be followed with the electroscope, the patient lying near the instrument which was at a fixed distance from the tip of the xyphoid process of the sternum. It appeared that, while there was excretion, it was relatively low and that less than 10% of the sodium was eliminated within the period of measurable activity (about 3 to 4 days). The quantity of sodium ingested (up to 53 mce.) produced no observable clinical effects.

Very interesting observations were early made in America on the circulation of sodium in humans (Hamilton and Stone 1937, Smith and Quimby 1944). For example, the radioactivity of the hand of the subject was observed after ingestion of Na24, the patient holding a Geiger–Müller counter in a lead box to screen off the effects of the rest of the body. Radio-sodium prepared by deuteron bombardment of 1 to 2 gm. of crystalline NaCl was dissolved in, say, 50 to 100 cm^3 of water, and then the appropriate amount of the diluted liquid was swallowed. Readings of the counter were made every few minutes. The kind of results obtained are shown in Figure 5, only some 80 to 200 μce. being required for

satisfactory results. The interpretation of such curves is difficult. The radioactivity of the hand may be used as an indicator of 'absorption', but is not necessarily a quantitative index of the rate at which the radio-sodium passes from the digestive tract into the blood stream. Moreover, the hand is a complex structure. If the rate of diffusion is rapid in the hand the counter will in effect measure the rate of change in activity of the incoming blood stream. If the rate of diffusion be low in comparison with the rate of absorption, the initial portion of the curve will be flatter and the interval of time before the maximum rise will be longer. Alteration of the calibre of the blood vessels will change the total volume of blood in the hand, and it was in fact observed that one subject after walking up four flights of stairs from the ground floor to the place where the tests were made showed a definite but brief increase in the activity of the hand!

Very little seems to be known about the quantity of sodium in the body or of its variation, but it is usually thought that practically all the sodium is contained in the body fluids, which comprise about 20% of the total body weight, and that diffusion is rapid. Even less is known of the effects of the total quantity of sodium in the body on the rate of absorption. In one case it appeared that absorption began in a few minutes, even when NaCl was ingested, and that equilibrium would be complete in some 3 to 10 hours, though in other patients the time required to reach equilibrium may be much longer.

Many studies have been made of the movement of sodium in the body and across cell membranes, Visscher *et al.* (1944) studying, for example, the interchange of sodium ions from the intestinal lumen into the blood. Flexner and his colleagues (1939, 1940, 1941a, b, c, d) have studied the transfer of sodium to the foetus in a number of species of animals including man, and find that in general the transfer is remarkably rapid, increasing greatly as gestation proceeds.

More recently extremely interesting observations have been made by Smith and Quimby (1944, 1945) who used radioactive sodium in studies of the circulation in patients with peripheral vascular disease. In these experiments (now comprising more than 200 patients) a few cubic centimetres of normal saline containing radio-sodium were introduced intravenously and the circulation of the blood traced by following the sodium it carried. Since the viability of an extremity depends upon its blood supply, it is evident that any information on this subject is useful in diagnosis, prognosis, and evaluation of therapy of vascular diseases. In practice the observations have been helpful in demonstrating the competency of the collateral circulation when the main arteries were partially or totally occluded. The method is simple, objective and does not require hospitalization. A Geiger–Müller tube shielded with $\frac{1}{2}$ inch of lead except for an aluminium window is placed against the region to be studied as the patient lies comfortably in bed. Usually the study is started with the window of the counter against the distal plantar metatarsal region of the foot. The material is injected into an antecubital fossa vein and the arm-to-foot circulation time established. The dose is usually about 100 microcuries in 5 cm^3 to 12 cm^3 of solution.

The arrival of radioactive substance in the tissues near the counter is indicated by a definite increase in counting rate. In normal subjects and in those having many vascular diseases the increase is marked and abrupt, but it is sometimes difficult to fix the exact time of arrival because of the background radiation. If the counter be left in position against the foot and the observations made minute by minute it is found that the counting rate increases steadily and markedly

for some time, the manner of 'build-up' of radioactive equilibrium depending on the rate of delivery of the isotope and on the relation between intravascular and extravascular space. The shape of the 'build-up' curves gives valuable clinical information relative to the circulation. Counts may be taken at other positions such as calf, popliteal region, or thigh. Comparison of counts at the same level in two legs is useful in determining the relative extent of the disease.

'Circulation time', that is the time interval at which a definite increase in activity begins, is extremely variable in different patients and pathological conditions, and ranges in Quimby's observations from 15 to 105 seconds, the former in a young man with pulse 100 suffering from scleroderma and the latter in an elderly arteriosclerotic. More valuable than 'circulation time', however, is the curve of build-up of equilibrium. In Figure 6 (*a*), (*b*) we give

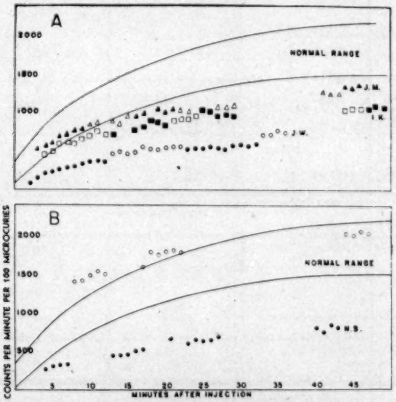


Figure 6. *After* Smith and Quimby (1945).

examples of observations obtained on patients suffering from frostbite and immersion foot. Fuller details may be found in the original paper (Smith and Quimby 1945). Two subjects, J.M. and J.W., suffered from immersion foot following exposure to water and cold while in open boats in the North Atlantic a year prior to the tests. The arteries of J.W. were palpable in the normal way though he complained of pain. I.K.'s foot was frozen in a blizzard nine years previous to the observation. N.S. (Figure 6 (*b*)) suffered from immersion foot following exposure in the South Indian Ocean in winter and had undergone a left lumbar alcohol block for pain. The resultant release of vasco-constriction is demonstrated by the very high sodium count on this side. The untreated right side shows the low count associated with immersion foot.

These results demonstrate the possibility of using the radioactive tracer method to determine the advisability of persisting in conservative as compared with more radical surgery and in deciding site of amputation.

More recently, studies of circulation by the use of Na24 have been undertaken in this country (Wright *et al.* 1948, Osborn and Wright 1949). Semi-automatic

methods of recording have been employed, and there can be no doubt that radioactive methods of studying the transport of materials in the circulation have considerable physiological and clinical utility.

We now turn to the possible therapeutic use of Na24, and again Mrs. Quimby and her collaborators have been responsible for many interesting investigations (Evans and Quimby 1946).

Radioactive sodium does not concentrate significantly in any particular organ but is rapidly distributed throughout the extracellular fluids of the body. Sodium is a normal constituent of body fluids, so that when the radioactive isotope is

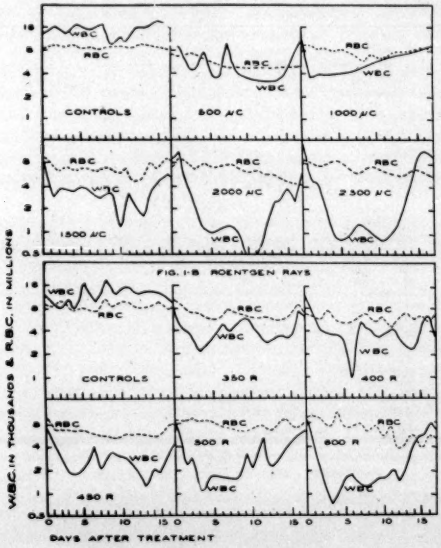


Figure 7. *After* Evans and Quimby (1946).

administered in a small amount of isotonic saline, no physiological disturbance should result except such as is produced by the effects of its radiation. Na24 might therefore be expected to be valuable as a substitute or supplement to other types of irradiation in the treatment of leukaemia and allied diseases in which whole-body irradiation is indicated. Before carrying out work on humans it seemed desirable to study effects on mice, and Evans and Quimby (1946) therefore carried out experiments on the comparison of the effects of whole-body x-irradiaton and the effects of Na24 on normal and leukaemic mice. The x-rays were generated at 184 kv. and used with a filtration of 4 mm. Cu + 2 mm. Al at a distance of 40 cm., using 10 r/min. Since two-thirds of the sodium radiation is given in the first day, 70% of the x-rays were also given during the first day and the remainder 24 hours later. Figure 7 shows the comparison and demonstrates the similarity of the

effects produced by radio-sodium and x-rays. Not only do Na24 and x-rays have similar effects in total white and red blood cell counts but they also produce similar changes in differential counts.

Interesting experiments were carried out on elimination by measuring with a Geiger counter the gamma-ray activity at various intervals over a period of 72 hours following injection. Figure 8 shows that the activity of the animals decreased more rapidly than expected from the known rate of decay of Na24, particularly marked effects being observed with a nursing mother known to eliminate sodium in the milk. The decrease curve is in all cases approximately exponential and may be regarded as indicating a mean life of some 14·3 hours in place of the normal mean life of 21·1 hours. This elimination of course decreases

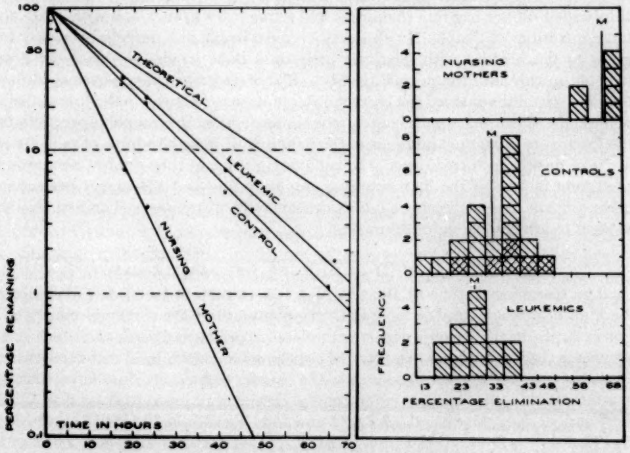


Figure 8. *After* Evans and Quimby (1946).

the 'integral dose' to the animal in this ratio. The experiments demonstrate clearly that the physiological changes produced by the beta- and gamma-rays of Na24 are similar to those attributable to whole-body x-radiation, in fact a quantitative equivalence exists between 10 μc/gm. and 100 r. of x-rays. It must be emphasized, for reasons explained later, that this equivalence would not be expected to hold for man.

Mice with enlarged nodes and extremely high leukocyte counts were found especially sensitive to radio-sodium, the marked response being apparently due to radio-sensitivity of abnormal leukocytes rather than any selective concentration of the material in lymph nodes. Somewhat different results have been obtained by Thygesen, Videboek and Villaume (1944), who have applied Na24 in the treatment of chronic leukaemia. They report favourable response in seven cases of chronic lymphatic leukaemia, but their results extend over a very short time (150 days). Lindgren (1944) found that P32 gave a more satisfactory response.

Of particular clinical interest are the studies of Fox and Keston (1945) of the mechanism of shock from burns and trauma in mice. The sodium content of injured skin is greatly increased. They have also noted a great influx of sodium into burned human tissue as determined by *in vivo* measurements at the bedside.

A few observations have been made using the long-lived Na22 by Reaser and Burch (1946), who have studied sodium metabolism in congestive heart failure; Cuthbertson and Greenberg (1945) have also employed Na24 in studies of dietary chloride deficiency.

Radio-sodium is also a useful source of radiation for more localized therapy. For example, Wallace, Walton and Sinclair (1949) have developed a technique for the treatment of multiple tumours of the bladder by introducing a thin-walled rubber bag into that organ and filling it for a few hours with relatively large quantities of Na24. In this way an even irradiation, largely beta, may be given to the bladder with relatively little total body irradiation since the dose falls off quickly into the normal tissues. Elaborate preliminary physical studies over many months enabled the correct order of dose to be given to the first patient, while at the same time remote control techniques reduced to reasonable proportions the hazards to personnel inherent in the handling of some hundreds of millicuries of Na24 during each treatment. In such circumstances the very low absorption coefficient in lead of the high energy gamma-ray line of Na24 makes protection quite difficult, but with care the difficulties may be overcome. The first results of such treatment are quite promising.

§6. THERAPEUTIC SOURCES OF GAMMA-RAYS

The foregoing outline of the use of a few important isotopes will indicate the kind of possible applications, but very many others are currently employed. For example, in the writer's department alone, experimental work and sometimes associated clinical work has been carried out using radioactive gold in the treatment of ascites, bromine in place of sodium in the bladder techniques, di-iodofluorescein in attempts to localize brain tumours, and preliminary physical work on S35, C14, Sr89 and 90, K42, Ga67 and 72, Ir192 and 194, Ta182, Tl 204, Sb122 and 124, Hg197 and 203, Sc46, Ca45 and 47, Cd115, Cs134, Y90 and 91, Cu64, Zn65 and As74. In addition to these specific problems there are a number of more general applications of gamma- and beta-ray emitters which we might mention.

We have already indicated that by neutron irradiation of stable elements in the pile and by the use of fission products very powerful sources of gamma-rays are now available. No final choice of source can be made, for it must clearly depend upon the therapeutic technique contemplated. Some of the considerations involved in the choice of β- and γ-ray sources have been discussed by the writer elsewhere (Mayneord 1946, 1950). For external powerful sources to be used in 'telecurie' units we require a long life, 'pure', preferably homogeneous high energy gamma-rays, free from primary beta-rays. The precise energy is of little importance biologically, but high quantum energy ensures high dosage-rate per disintegration and also improves slightly the penetration of the beam in tissues. High specific activity is of primary importance, since the best techniques make use of remotely controlled sources whose weight and linear dimensions should be small to ensure ease of handling and reasonably sharp edges to irradiation fields. The smaller the source, too, the easier the provision

of adequate protection for patient and staff, an important matter since the next five years may well see experimental therapeutic units containing 1,000 curies of powerful gamma-ray emitters.

Gamma-ray sources for interstitial or surface therapy require to have rather different characteristics. Small rod-like sources are required (of the order of 0·5 to 5 mc.) and freedom from beta-rays is of enhanced importance in order to avoid local energy absorption and consequent necrosis. The homogeneity of the radiation or its precise quantum energy matter relatively little since dose is determined largely by geometrical rather than physical considerations. The inverse square law rather than absorption coefficient determines dose distribution. There is need of both long- and short-lived substances, the former to be left in tissues only a limited time and removed at operation and the latter to be left indefinitely to 'die' *in situ*.

A large number of possibilities arise, but so far little systematic work has been carried out on this subject, which requires intensive effort. It seems that no fundamental improvement is to be expected over the present 'radium needle' techniques, but in detail important technical advantages might be gained. We might illustrate these advantages by considering the possible use of Co60, the most probable 'substitute for radium'. It has already been much used in physical laboratories, and in the writer's experience is an extremely valuable standard and source of nearly monochromatic gamma-radiation. Its physical characteristics are well known, having a half-life of 5·3 years and emitting two gamma-rays in cascade of 1·16 and 1·32 Mev. respectively. It may for many purposes be regarded as emitting monochromatic radiation of energy 1·2 Mev. The absorption coefficients of these radiations in a number of materials have been measured (Mayneord and Cipriani 1947) and shown to be in accord with theoretical predictions of Klein–Nishina and Fowler–Hulme theories of Compton and photoelectric absorptions respectively.

Co60 may be produced by a simple (n, γ) reaction on Co59 which constitutes the whole of normal metallic cobalt. Owing to this fortunate fact and the relatively large cross section it may be produced in large quantities, and in high neutron flux yields high specific activity. The advantages we might expect are as follows. The gamma-rays from cobalt are simpler than those from RaB and C, which contain components ranging from very low energies up to 2·2 Mev. We do not need, therefore, to use a thick metal sheath to remove the easily absorbed γ-ray components. The old 'filtration' problem of radium therapy disappears. In addition, Co60 decays to Ni60 (which is of course the real source of the gamma-rays) and there is no complex chain of breakdown products to complicate dosimetry. The problems of producing equilibrium conditions are transferred to the irradiation in the pile. Moreover, there are no troublesome gaseous products, a continual danger with radium containers. The cobalt may be used in the form of a thin metal rod or wire, thus avoiding the dangers and complex problems of filling small quantities of fine radioactive powders into hollow containers. Again the cobalt emits only beta-rays of maximum energy 0·3 Mev. with its gamma-rays, as opposed to the troublesome energetic beta-rays of RaC, and generally it is easily calculated that whereas the beta-rays in equilibrium with radium and its products amount to some 70% of the gamma-ray energy emitted, for cobalt this beta-radiation is only about 5% of the gamma-ray energy emission. There is therefore much less local high dose. The atomic number of cobalt

also ensures a minimum electronic emission under its own gamma-rays, as opposed to the very large emission from the heavy platinum containers necessarily used to contain radium in view of the high energy of the beta-rays. The cobalt source must, however, be enclosed for two reasons, first to remove the beta-radiation which, though small, still implies high local energy absorption and, therefore, tissue destruction near an embedded source, and secondly because it has been shown that cobalt metal is absorbed from subcutaneous sites and can be found in animals into which cobalt has been implanted, in adjacent tissues, liver, kidneys and excreta (Nickson, Lamerton and Mayneord 1949). Perhaps wire sheathed in gold tube or in stainless steel containers may prove satisfactory.

The disadvantages of cobalt seem to be its relatively short life, which implies a loss of activity of about 1% per month, which may prove troublesome in clinical practice, and the need to use high neutron flux to obtain the specific activities required. Much of the advantage is lost if sources of diameter greater than about 1 mm. are used. Fortunately, with the advent of more powerful piles this problem seems to be solved, and it is likely that, particularly if heavy water piles may be used for activation, very high specific activities of the order of curies/gm. will be possible which enable flexible wire sources to be made having activities of the order of 1 mc. per cm. length.

Detailed study to be found elsewhere (Mayneord 1950) shows that in order to rival modern x-ray equipment in output and penetration, very strong sources would be required, of the order of 1,000 curies. Such a source would be the equivalent in gamma-ray output of 1·63 kg. of radium, and is clearly a very dangerous as well as a powerful weapon. It could probably be designed and operated safely, particularly if metallic uranium be used as its protective shield, but it remains to be seen whether the very grave technical difficulties could be justified by the somewhat doubtful advantages as a therapeutic source of radiation.

Cobalt is by no means the only possible source of this character, and a good deal of preliminary work has been done on iridium (Freundlich, Haybittle and Quick 1950), which also promises to be useful therapeutically and is already used fairly extensively as a source of gamma-rays for industrial radiography.

§7. EXTERNAL BETA-RAY SOURCES FOR THERAPEUTIC USE

The use of beta-ray sources therapeutically on the skin has frequently been looked upon with suspicion, yet there can be no doubt that in the treatment of superficial lesions they can be extremely useful. The physical difficulties hitherto have been the complex and unsatisfactory nature of most of the beta-ray sources available. The usual radium sources in very thin-walled containers are often very variable in output over their surfaces and suffer from mechanical fragility with its consequent danger of leaks of radon or even radium. Moreover, the admixture of powerful gamma-radiation makes any strict beta-ray therapy, limited in depth to the range of the particles, quite impossible. Yet the extreme portability as compared with its rivals (low-voltage x-ray equipment), the high localized dosage-rates and consequent ability to treat large numbers of patients and the ease with which surrounding tissues may be shielded from irradiation, offer attractive possibilites.

The use of artificial beta-ray sources promises to give a considerable impetus to this form of therapy. We now have available a number of substances, some of them fission products, P32, Y91, Sr89, and many others which emit pure beta-rays

free from gamma-rays. The choice of therapeutic source is somewhat difficult. One approach is to use activated foils or to incorporate such substances into plastics (Raper 1947) with the idea of moulding the plastic to the surface to be treated and subsequently activating it in the pile. The question is not quite so easy as there are many chemical difficulties, while other elements, such as sodium, in the plastics may be strongly activated and emit powerful gamma-rays. Moreover many materials disintegrate under intense neutron bombardment, but it seems likely that success will be attained and very useful sources made available. In unpublished investigations Sinclair, in the writer's Department, has used such sources of P32 and preliminary clinical work has been carried out. For example, moulded applicators for the treatment of some diseases of the eye have been developed, and it seems that the technique is a very promising one. Of course, again, specific activity is important, activities of the order of 1 mc/cm^2 being required in therapeutic practice. Suitable choice of active material should enable lesions at different depths to be treated, using beta-ray emitters of varying energies.

Recently experiments have been performed using beta-ray sources consisting of an element of long life followed by a product of short life and therefore emitting high energy beta-rays, thus obtaining long-lived sources emitting energetic beta-particles (Freundlich 1949). The precise detail of the future therapeutic sourcesremains to be developed, but physical studies are in progress which bid fair to indicate the optimum physical and geometrical conditions of use. There is little doubt that the method is likely to prove useful. A beginning has, for example, been made in the United States (Low-Beer 1946, 1947, 1950) by the use of blotting paper dipped in solutions of P32 and cut to the shape of the lesion to be treated. Such a technique is evidently open to criticism on physical grounds, and must introduce considerable uncertainties of dosage, but it represents an interesting beginning and provided useful clinical experience.

§ 8. DOSIMETRY OF ARTIFICIAL RADIOACTIVE SUBSTANCES

The medical applications of artificial radioactive substances have brought in their train a number of interesting dosimetric problems which have so far been inadequately treated yet are of urgent practical importance. From the medical point of view the main effort must be directed at bridging the gap between the dosimetry of beams of radiation, as used in the now classical techniques of radiation therapy, and the relatively new field of dosimetry of radiations emitted by radioactive substances extended throughout the body after injection or ingestion. In this way alone can the amounts of the new materials which may be used, or may require to be used in order to produce significant biological effects, be estimated. We must use the accumulated experience of required dosage levels in radiation therapy on the basis of estimation. The mathematical results required are in the main already available, for the study of the amount and distribution of radiation around extended radioactive sources has been a favourite subject for many years (Sievert 1932, Mayneord 1932, 1945, Meyer and Schweidler 1927). The new dosimetry illustrated admirably the advantages of an absolute energy system of units of dose, for only in terms of energy absorption, measured essentially as ergs/gm. of tissue, can any sure basis be found for the new dosimetry. Moreover, the recently developed theory of total energy absorption from beams and point sources of radiation may be adapted by virtue of single reciprocity

relationship to the dosimetry of artificial radioactive substances. In many cases all that was required was to re-define the symbols of existing investigations. A common energy system of dosage also enables a general theory of dosimetry to be sketched for a wide range of ionizing radiation, including gamma-rays, beta-rays, positrons, neutrons and heavy particles, It must not be forgotten in applying such a theory that there is now convincing evidence that the same amount of energy absorbed, for example, from a beam of alpha-rays may produce a quantitatively different effect from that brought about by the same absorbed energy from gamma-rays, the difference resulting from differences in specific ionization, that is spatial distribution of ions relative to the effective biological units. These biological considerations are best introduced separately, and subsequently to the purely physical considerations of energy absorption.

(i) *Energy Absorption and the Röntgen*

It is now well recognized that the röntgen is essentially a unit of energy absorption in a given mass of air, and it is easily shown that one röntgen delivered to 1 gm. of air results in the absorption of approximately 83 ergs (Mayneord 1940, 1944, 1945, Mayneord and Clarkson 1944). In soft tissues the corresponding energy absorption is approximately 95 ergs owing to the presence of hydrogen with twice the normal mass absorption coefficient per gm. for light elements. The unit of 83 ergs has been called by the writer a 'gramme-röntgen' (gm.r.) and has been used as a unit of absorption in the comparison of the general constitutional effects of different types of therapeutic procedures. The energy absorption per röntgen per gramme of other materials of biological interest has been studied elsewhere up to about 1 MeV. by the writer (Mayneord 1937) and by Spiers (1946). The extension of these calculations to higher energies has recently been carried out and leads to interesting and useful results, which will be discussed below (Mayneord 1950, Failla 1945, Lauritsen 1933, Streib, Fowler and Lauritsen 1941).

It is of importance to calculate the energy flux required to record one röntgen in beams of gamma-ray quanta of various energies. The fraction of the incident energy 'converted' into the energy of electronic motion (in which form alone it is assumed to be available for producing biological effects near the point of absorption) may be written $(\sigma_a+\tau_a+\pi_a)N\,dm$, where σ_a is the Compton recoil coefficient associated with production of recoil electrons, τ_a the photoelectric absorption coefficient corrected for the binding energy of electrons and π_a a new coefficient we introduced as representing the real energy absorption by pair formation. Thus if a quantum of energy $h\nu$ be absorbed, the energy available biologically at pair formation is approximately $h\nu-2mc^2$, where m is the rest mass of the electron and c the velocity of light, it being assumed that annihilation only occurs after the electrons have been brought to rest. In the subsequent annihilation of positive and negative electrons, $2mc^2$ is of course recovered, but as gamma-rays whose utilization may be inefficient and spread throughout a large volume. This annihilation radiation must be treated as a separate gamma-ray problem.

We set $\pi_a=\pi\{1-2/\alpha\}$ where, in the usual notation, $\alpha=h\nu/mc^2$ and π is the pair formation coefficient, i.e. the cross section for pair formation per atom. The coefficients may then be calculated for various energies of quanta and materials as follows: σ_a for the Klein–Nishina formulae, τ for the Hulme–Fowler

theory and π for Heitler's calculations. In this way we may calculate the energy flux per röntgen up to 100 Mev. (Figure 9), whence it will be seen that the value remains roughly constant at about 3,000 ergs/cm² r. from 0·1 to 1·0 Mev., followed by a slow rise to about twice this value to 10 Mev. followed by a second fairly constant region from 10 to 100 Mev., though it must be remembered that above some 30 Mev. the values are to be regarded as first approximations only. In a more extended report (Mayneord and Martin 1950) the theoretical investigations are detailed to other media and conditions. The calculations are of interest in the measurement of high energy quanta since they give us the beginnings of a theory of the ionization chamber in the region above 1 Mev. In all these considerations the work of Gray (1937) is fundamental.

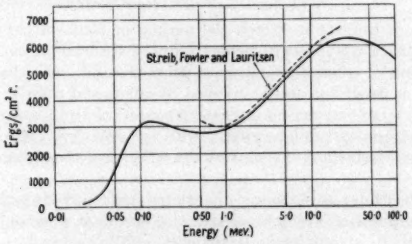


Figure 9. *After* Mayneord (1950).

(ii) *Dosage Rates due to Radioactive Sources*

Problems of dosimetry in medicine frequently fall into two categories, namely those concerned with units and absolute dosage-rates and those arising from the study of the distribution of radiation in a scattering medium such as the human body. The foregoing calculations enable certain useful generalizations to be arrived at concerning the dosage rates in r/hr. near radioactive sources.

Consider a point source of strength one curie (defined as the amount of material undergoing $3{\cdot}7 \times 10^{10}$ disintegrations per second) of any gamma-ray emitting substance, assuming at this stage, for simplicity, one quantum of energy E_{Mev} to be emitted per disintegration. Then the intensity of radiation at 1 cm. distance will clearly be

$$I = \frac{1}{4\pi}\{3{\cdot}70 \times 10^{10} \times 3600 \times E_{\text{Mev}} \times 1{\cdot}59 \times 10^{-6}\} = 1{\cdot}685\, E_{\text{Mev}} \times 10^{7}\ \text{ergs/cm}^2\,\text{hr}.$$

In practice the most convenient unit is the dosage rate in r/hr. from a millicurie source:

D = dosage rate in r/mc. hr. $= 1{\cdot}685 \times 10^{4} \times E_{\text{Mev}}/E_{\text{f}}$ where E_{f} = energy flux per röntgen already calculated. Making use of this relationship we calculate the dosage rates to be expected from point sources emitting gamma-rays of different energies. Owing to the approximately constant energy flux per röntgen from 0·1 to 1 Mev. the dosage rate is in this region approximately proportional to quantum energy (r/mc. hr. $\simeq 5{\cdot}6 \times E_{\text{Mev}}$), but at higher energies the dosage rate increases less rapidly. The dosage rate for actual radio-isotopes may clearly be obtained by superposition of the dosage rates of individual quanta corrected for

their respective chances of emission. The only published experimental evidence at present available as to the accuracy of these calculations seems to be that for radium sources (Gray 1937, Mayneord and Roberts 1937, Taylor and Singer 1940). Such investigations lead to a value of 8·3 r/mc. hr. for a filtration of 0·5 mm. platinum or 9·0 r/mc. hr. for the gamma-rays if unfiltered. The calculated value is 9·2 r/mc. hr. There is evidently considerable need of calorimetric experiments on a number of isotopes as the results are of fundamental importance. Meanwhile the curve gives a useful guide to expected dosage-rates. In practice we also made use of it in the measurement of gamma-ray emitters since a measurement with an air-wall chamber in röntgens now permits of an estimate of strength of source in millicuries.

(iii) *Dosimetry of Beta-Ray Emitters*

It may be of interest to sketch the outline of some of the problems now confronting medical physicists in the dosimetry of artificial radioactive substances and the methods of solution which have proved useful. We shall not consider the problems in detail but outline method of attack and main results.

The dosimetry of beta-ray (negatron) emitters is relatively simple and has been considered by Hertz and collaborators (1940) for iodine, and more generally in a useful paper by Marinelli (1942) and by the writer in a recent report (Mayneord 1950).

Consider then a negatron emitter distributed uniformly throughout the tissues with a concentration of 1 μc/gm. of tissue. The total number N of particles emitted when 1 μc. is completely destroyed (1 μcd.) is given by

$$N = 3{\cdot}70 \times 10^4 \times T/0{\cdot}693 = 5{\cdot}34 \times 10^4 T,$$

where T is the half-life in seconds. In a medium of extent large compared with the range of the beta-particles and uniformly filled with radioactive material the energy absorbed per gm. of tissues is equal to the energy emission per gm. or $E = 5{\cdot}34 \times 10^4 T \times E_{\text{Mev}}$ Mev/gm., where E_{Mev} is the average energy of the beta-rays in Mev. and is usually of the order of one-third the maximum energy.

We know, however, that when 1 gm. of air receives 1 r. the energy absorbed is 83 ergs, and we may, by assuming equality of air and soft tissues, calculate the equivalent röntgens on this basis. A little more accurately we may note that in 1 gm. of water 1 r. of gamma-rays corresponds to approximately 95 ergs. 1 μcd/gm. of tissue therefore corresponds in energy absorption to $5{\cdot}34 \times 10^4 T E_{\text{Mev}}/5{\cdot}22 \times 10^7 = 1{\cdot}02 T E_{\text{Mev}} \times 10^{-3}$ r. if T be in seconds, or $88\, T E_{\text{Mev}}$ r. if T be in days. Alternatively, 1 μcd. $\simeq 88 \bar{E} T$ gm.r.

The integral dose per day for 1 μc. initially injected will be given by: Dose per day per μc. = Dose per μcd. $[1 - e^{-0{\cdot}693/T}]$. For example, for P32 with $T = 14{\cdot}3$ days and $\bar{E} = 0{\cdot}70$ Mev., 1 μcd. = 880 gm.r.

The integral dose per day for 1 μc. of P32 at the commencement of the period is: Integral dose $= 880 \times 0{\cdot}0485 = 42{\cdot}7$ gm.r.

If the whole of the μc. be contained in 1 gm. of tissue the 'equivalent dose' is 42·7 r/day. A tissue whose radioactive content is 1 μc/gm. of P32 will therefore be subjected to energy absorption per day equal to that due to 42·7 r. The general relationships are best seen in terms of gm.r. and ergs rather than röntgens since 1 μcd. must give rise to a given emission and absorption of energy independent of the distribution of radioactive material. Details for various isotopes may be found elsewhere (Marinelli, Brinckerhoff and Hine 1947).

It may be noted that we now have considerable experience of the dosimetry of whole-body x- and gamma-radiation and know, for example, that 10–30 Mg.r. of x-rays in six weeks would be likely to produce profound effects on the blood picture. This integral dose would correspond to 11·4 to 34 mcd. of P32. It is therefore not surprising that in the treatment of leukaemia with P32 it is observed that similar changes are produced by doses of the order of 5–20 mcd. in a few weeks. Clinical experience therefore bears out the correlation of x-ray and beta-ray doses predicted theoretically on a pure energy basis, at least as to order of magnitude.

This problem is fundamental in the medical applications of radioactive isotopes, and there appears to be evidence that in some instances the problem is much more complex than indicated by these simple energy correlations. Here there is much need of research.

(iv) *Dosimetry of Gamma-Ray Emitters*

The solution to the problem of negatron emitters is fairly obvious and elementary, but that of gamma-ray emitters is more difficult mathematically. We may however make a modest beginning, restricting ourselves to the simplest geometrical conditions, as follows. It is well known that the apparent absorption of gamma-rays in large blocks of tissue in which a gamma-ray source is buried is very small and may frequently be neglected. That is, only inverse square law considerations need be used to attain first approximations to dose distributions. We may therefore write down the dose distribution for the simplest possible case, namely a sphere filled uniformly with radioactive material. At a distance c from the centre the dose rate is given by

$$Dc = k\pi\rho\left\{2a + \frac{a^2-c^2}{c}\log\frac{a+c}{a-c}\right\},$$

where ρ is the radioactive density, a the radius of the sphere and k the constant previously calculated. If k is in r/mc. hr. then ρ must be in mc/cm³ of tissues. At the centre of the sphere the dose rate is obviously $D_{\text{centre}} = 4\pi k\rho a$. Knowing k for the various isotopes we may therefore write down the dosage rates to be expected for given uniform concentrations of material. In the case of substances such as Na24 or I131, emitting both beta- and gamma-rays, by simple addition the final energy absorption may be found. It is very important to observe one important difference between beta- or gamma-rays, corresponding to their very different ranges. For gamma-rays the dosage rate is roughly proportional to the linear dimensions of the mass of tissues containing the material, whereas for beta-rays, whose mean ranges are frequently of the order of a few millimetres, the dose is independent of size. For small masses of tissues therefore the beta-ray effects tend to predominate, but for large volumes of the order of the size of human beings the gamma-ray effects have outweighed them. This is important in comparing the effects on mice and men! For the beta-ray dose it may be permissible to expect rough equality of effect per μcd. per gm. of tissue, but for gamma-ray emitters this is no longer so. When, of course, the order of size considered is again reduced we find exactly similar problems in comparing beta- and alpha-rays, but we must not stay to discuss such detail here.

Passing from the spherical masses hitherto considered we may extend the calculations to other shapes such as ellipsoids and cylinders, making use of the previously developed mathematical theory of radiation fields. The theory is in

essence that of integral dose, since we may make use of important reciprocal relationships between the energy absorbed by any mass of tissues from a point source of radiation and the dose rate at the position of its point source if the absorbing mass be itself considered as uniformly filled with radioactive material. Detailed calculations relating to various masses of tissue and the effects of absorption corrections on the distributions may be found in papers by the writer and others (Mayneord 1950).

Account may also be taken of unequal distributions of radioactive material, but in general such calculations are tedious and involve numerical or graphical integration.

Applying these methods, sufficiently accurate estimates may frequently be made of the energy absorption per gramme of tissues even in the complex geometrical conditions of the human body, provided of course the actual distribution of radioactive material be known, or some simple assumption may reasonably be made about it. The mathematical theory of this dosimetry is, in fact, well developed provided the assumption of uniform density of radioactive material be made, but unfortunately this is rarely true in practice. It appears that, as so often before, the physical theory gives an excellent general guide, but the circumstances of each individual patient are so different as to call for detailed examination. The present situation seems to be that we now require quick reliable methods of finding the distribution of radioactive material in each patient before we can make much progress in our dosimetry. The writer has indicated elsewhere (Mayneord and Belcher 1950) how by the use of well canalized scintillation counting systems, such problems may be tackled, but the interpretation of results will always call for much caution.

(v) *Dosimetry of Positron Emitters*

The theory of positron emitters may also be treated in a similar way, calculating both the energy loss from ionization by the particles and the number of quanta of annihilation radiation from the number of positrons brought to rest per second and using $k = 3{\cdot}0$ r/mc. hr. for the gamma-rays of energy 0·51 Mev. subsequently emitted.

The problem of gamma-ray fields due to positrons appears again in considering the secondary effects of annihilation radiation due to positrons arising from the absorption of high energy quanta. For example, in the protection of personnel against very high energy radiations we may calculate that the energy emitted as annihilation radiation in an element of mass dm per röntgen delivered to it is

$$E = \frac{167{\cdot}4\pi_a}{(\sigma_a + \pi_a)} \frac{mc^2 dm}{(h\nu - 2mc^2)} \text{ ergs.}$$

Knowing the energy flux per röntgen for annihilation radiation (2,820 ergs/cm²r.) we may now calculate the dose-rate due to annihilation radiation coming from a uniformly irradiated mass, by integrations analogous to those described. It is found that the effect reaches a maximum in the region of 5–10 Mev. as might perhaps have been anticipated since for lower energies pair formation accounts for only a small part of the total energy absorption, while at high energies the mass equivalent of the electron is small compared with the initial kinetic energy of emission of the electrons. Detailed calculation shows that for the human body, subjected to irradiation by quanta whose energy lies in the

region of 5–30 MEV., it is unlikely that annihilation radiation will account for more than some 2% of the energy absorption at any given place.

(vi) *Fast and Slow Neutrons*

Of more physical interest than these largely geometrical studies are the attempts to correlate the dosimetry of beams of fast and slow neutrons with absolute energy considerations. When beams of fast neutrons enter the body it seems that most of the initial biological effect is due to recoil nuclei, particularly to protons, and to a less extent recoil carbon nuclei. Mitchell (1947a, b, c) has considered the problem in detail and given very valuable estimates of both absolute energy absorption and also the distribution of that absorption at a depth in soft tissues. Tait (1948, 1949, 1950) has later improved the methods of calculation of energy distribution and enabled more accurate estimates of probable biological damage to be obtained. We cannot discuss the calculations in detail here but they have great significance in, for example, calculating 'tolerances' of high energy neutrons. When slowed down these neutrons at thermal energies are largely captured by protons to form deuterons with the emission of a quantum of 2·2 MEV. gamma-rays. The theory of these important gamma-ray fields may evidently be treated as a special case of that outlined above. In detailed calculation of energy absorption due to neutrons account must also be taken of other nuclear reactions of thermal neutrons, particularly the (n, p) reaction with nitrogen of the protein.

§ 9. MEDICAL USE OF HIGH ENERGY QUANTA AND PARTICLES

In addition to the gift of artificial radioactive isotopes, nuclear physics has presented to medicine particles of extremely high energies. These particles may be accelerated in a variety of different ways, as, for example, by the use of a Van de Graaff generator, cyclotron, betatron, synchrotron, frequency modulated cyclotron, linear accelerator, and other instruments in process of development. It is quite beyond the scope of this article to discuss the physical detail of these instruments, which have in any case been severally discussed in the literature (Institute of Physics 1950), and we will confine our attention to a few of the possibilities as seen from a medical point of view, describing tentative attempts so far made to utilize some of these generators in medical practice.

The cyclotron itself, being able only to accelerate heavy particles, has not so far been used directly in medicine, though, of course, its influence on the development of high energy accelerators has been very great. We may perhaps remark, however, that a beam of high energy protons would seem to be a very promising therapeutic weapon could it be applied. From the point of view of delivering accurately a large amount of energy to a depth in tissue it has advantages in that, owing to the small scattering of heavy particles, well-defined beams even at a depth should result, while in accordance with the Bragg curve a large absorption of energy would occur near the end of the range, so that by varying the initial energy of the particles lesions at different depths could be treated. The physics of this application has been worked out in some detail by Wilson (1946) and others, but so far, unfortunately, no machine has been available for medical research or clinical use. Protons of the order of 100–150 MEV. initial energy would be required.

§10. THERAPEUTIC USE OF HIGH SPEED NEUTRONS

The use of beams of fast neutrons in a way analogous to beams of X-rays has been much discussed and one important clinical investigation made by Stone and his colleagues (1940, 1942) into the matter. The beam of neutrons was produced by the bombardment of beryllium with deuterons at 60 MeV. The maximum energy of the neutrons was 16 MeV. but the precise energy distribution is not known. By the use of special applicators, consisting of paraffin wax tunnels having lead filters incorporated in them, the gamma-ray intensity was reduced to small proportions and a fairly well-defined beam of neutrons was obtained as shown by the sharpness of the edges of reaction fields on the skin of animals and patients.

The 'penetration' of the beam as indicated by a standard X-ray dosemeter was about the same as for a 200 kv. X-ray beam, but in view of the probable complexity of response of the measuring instrument to neutrons of varying energies and the probable varying significance at different depth of capture gamma-rays, the results are to be regarded with caution. One of the most complex fields of dosimetry facing the medical physicist in the next few years is this dosimetry of neutron beams in scattering hydrogenous media. The investigations by Gray and his co-workers (Gray 1937, 1943, Gray and Read 1939) have, however, now laid a much more secure basis for this work than was available to the American workers.

The results of irradiation by neutrons are similar to those of X-rays, no striking qualitative differences in biological or clinical effects being observed. Since only incurable disease was treated by the method, results as shown by statistical analysis of survival are rather depressing, and there are indications of serious late effects. Nevertheless, it is of importance that these experiments be followed by others on the clinical use of high energy neutrons. A consideration of the physical mechanism of biological effects as compared with X-rays leads to a possibility that differences might be observed. For example, the specific ionization along the tracks of recoil protons at different depths in tissue will be a function of the mean energy of the neutrons producing these protons, and therefore of depth below the skin. There may therefore be relative effects at a depth compared to those on the skin which might have advantages in the treatment of malignant disease. It is a matter of importance that there be available in this country a cyclotron for the continuation of therapeutic experiments with neutrons. It would, of course, also serve as an extremely valuable source of radiations for studies of protection at greatly reduced dosage levels, as well as fundamental studies on the relative biological efficiencies of energy absorbed at different specific ionizations along the tracks of the effective particles.

§11. HIGH ENERGY X-RAYS

For many years the tendency of X-ray therapy has been towards the use of higher voltages, the main purpose being to attain greater penetration of radiation, thus enabling lesions at greater depths to be treated effectively. At the same time much effort has been expended in the production, accurate alignment and application of small beams of radiation with accuracy to the patient. Moreover, fairly comprehensive studies of radiation distributions within the body resulting from these techniques have shown clearly the need for further development in this direction if the control of the distribution of energy is to be satisfactory.

The production by Kerst of the betatron and its subsequent development to accelerate electrons to energies up to 100 MEV. has placed in our hands an instrument of great interest which is now being applied in therapeutic practice.

At very high energies the betatron becomes a very costly and cumbersome instrument, so that the development of the beta-synchrotron, with its promise of smaller, lighter machines, is of medical importance. It must be realized that in medical use mobility and a high order of precision of application may be of greater importance than maximum energy attainable.

Since the quantum radiation produced by betatron and synchrotron is the same, for our purpose we may deal with both instruments together.

Two questions invariably arise in connection with a new therapeutic weapon of this kind: (i) is it likely that there are any qualitative differences to be expected in the biological effects of the new radiation, and (ii) what will be the distribution of absorbed energy in the human body subjected to it, for this geometrical factor may be of the greatest significance? We may now consider these questions from the physical point of view.

The mechanism of energy absorption of high energy quanta in human tissues will differ from that of ordinary X-rays in two important respects: (i) in the production above 1·1 MEV. of electron pairs, and (ii) in the production of nuclear disintegrations.

No entirely satisfactory studies of the very difficult problem of energy distribution in the continuous X-ray spectrum produced, say, from a tungents target by high energy electrons are available, but there is evidence that the distribution may be approximately represented by the predictions of Heitler. If, with this assumption, we apply also the Heitler theory of pair formation, it seems that this process is not of great importance ($<15\%$) below, say, 10 MEV., so that the absorption is still very largely by production of Compton recoil electrons in a medium of low atomic number. At some 30 MEV. the two processes account for equal energy absorption in air (Mayneord 1950). It is, therefore, unlikely that there will be large biological differences in efficiency per röntgen up to this order of energy, though there may be a little owing to the low specific ionization along the tracks of high energy electrons. The very limited clinical observation so far bears out this prediction. No qualitative differences have been observed in practice between the effects of 20 MEV. X-rays and those produced at 200 KV., but it appears that rather larger doses (of the order of 20%) are required to produce the same effect. The precise relationship is a little different for different biological reactions, such as effects on bacteria, greying of hair, inhibition of growth of organisms, induction of mutations. A detailed consideration of the biological and medical significance of nuclear reactions produced by high energy X-rays may be found in a previous report by the writer (Mayneord 1950).

The most probable photo-disintegration is usually a simple (γ, n) reaction resulting perhaps in a positron-emitting isotope of the irradiated material. Many other types of reaction occur, but the cross sections are low. Using a 30 MEV. synchrotron the writer and his colleagues have observed the production of radioactivity in animal tissues by X-rays produced from 17 MEV. to 24 MEV. (Mayneord, Martin and Layne 1949). The main disintegrations result in the production of positron-emitting O15 and C11. The activity of the tissues decrease at first with a half-life of about 2 minutes, due to the oxygen, and eventually with a half-life of about 20 minutes, due to the carbon. Detailed study shows that the rate of absorption of energy due to activation is only of the

order of 0·01% of that due to the more usual mechanisms of absorption, namely electron recoil and pair production. It is unlikely, therefore, that the production of radioactive isotopes in the tissues contributes materially to the biological effects of high energy x-rays in them.

We have recently observed the production of radioactivity in patients being treated with the synchrotron, and the phenomenon may be of value in showing the limits of irradiation and also even in measuring the dose received at a depth.

(i) *Dosimetry of High Energy X-Rays*

The measurement of x-rays produced at energies about 2 Mev. requires special consideration. It is not possible to discuss the problem in detail and we will limit ourselves to energies below 30 Mev. Further detail of both theory and experimental investigations made by the writer and his colleagues on both sides of the Atlantic will be found in the report to which reference has several times been made (Mayneord and Martin 1950). Briefly, the main difficulty arises from the long paths of secondary particles, particularly electrons, in the material of the wall of ionization chambers submitted to high energy quanta.

In materials of unit density these paths are roughly 1 cm. per 2 Mev. initial energy, so that to attain electronic equilibrium in a measuring chamber it is necessary to use very thick walls. These walls then contribute usually unknown amounts of scattered radiation, introduce serious absorption corrections and also make impossible the precise definition of the point at which the beam is measured. Figure 10 shows, for example, the variation of ionization current in a carbon

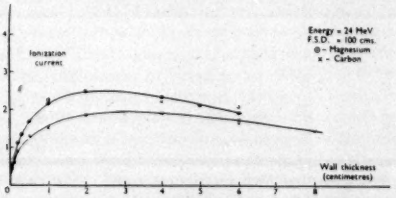


Figure 10. *After* Mayneord and Martin (1950).

chamber as its front wall thickness is increased in a beam of x-rays produced at 24 Mev. In practice the presence of electrons in the vicinity of a high energy accelerator usually still further complicates interpretation. It happens, however, that the definition of the röntgen requires electronic equilibrium in a chamber if the correct measuring conditions are to be fulfilled. The fact that measurements may be made in the region of 20 Mev. is illustrated by the fact that Mayneord and Martin in a comparison of measurements of output of betatrons, synchrotrons, linear accelerators and electrostatic generators in America and Great Britain agreed with Kerst's estimate to within 7% at 20 Mev. in the case of Kerst's betatron when somewhat arbitrary measuring conditions were imposed. The present position is that while approximate measurements can be made in röntgens, the definition of the conditions becomes increasingly arbitrary as the energy rises, and therefore, certainly beyond about 20–30 Mev., it is preferable to attempt an estimate of real energy absorption from the 'cavity' ionization measurements as first indicated by Bragg and Gray. We are again driven to an

absolute energy unit of 'dose', namely ergs/gm., as in the isotope and neutron dosimetry.

The build-up curve of Figure 11 illustrates one of the most important properties of these high energy quanta in clinical use. It is clear that the energy absorbed per unit mass in the tissues will increase with depth up to the maximum at a depth of the order of the mean range of the secondary electrons liberated in the tissues. The depth at which the maximum occurs in a medium of unit density is shown in Figure 12. We should expect, therefore, that in clinical practice very

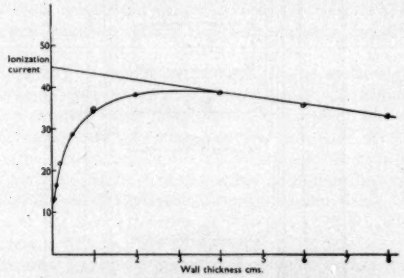


Figure 11. *After* Mayneord and Martin (1950).

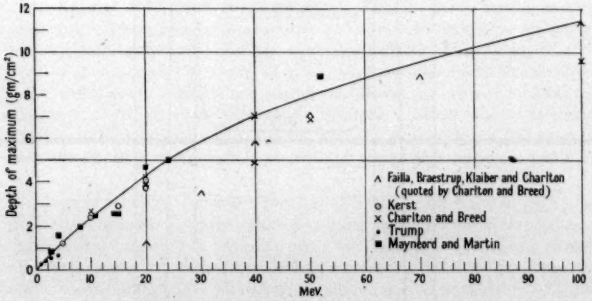


Figure 12. *After* Mayneord and Martin (1950).

little, if any, skin effect would be observed even for effective doses at a depth. This is a very important consideration, as skin effects have frequently been the limiting factor in clinical dosimetry. We note, too, that any effects there are should be looked for at the point of exit, not of entry, of the beam, since below the maximum the dose falls off very slowly with depth.

In our very limited clinical experience these predictions are borne out. The absence of surface skin effects, so that profound destructive changes are possible at a depth with little to show on the skin, emphasizes the importance of detailed preliminary physical studies of dose distributions in watery media. A number of

such investigations have been made, and we show in Figure 13 the 'isodose curves' obtained by Johns and his co-workers with a 22 Mev. betatron (Johns *et al.* 1950). Kerst's and our own measurements are in substantial agreement with the results shown.

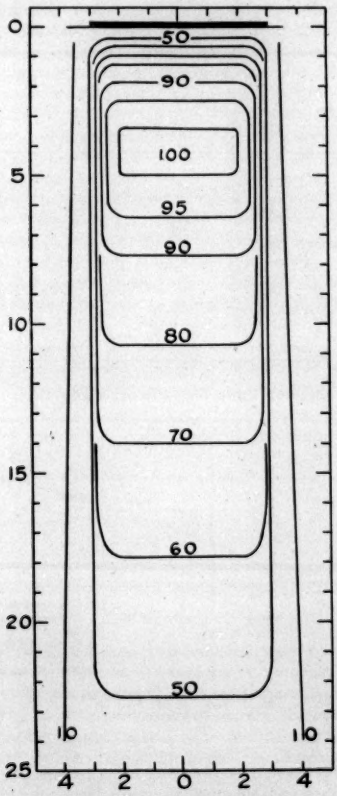


Figure 13. *After* Johns *et al.* (1950).

A physical factor which is of importance in therapeutic applications is the spatial distribution of the x-rays produced about the target. Whereas at low energies x-rays are emitted in all directions, as the energy of the impinging electrons increases so the energy is emitted more and more nearly in the original electronic direction, the x-radiation being confined in a solid angle of the order of mc^2/E. The precise distribution to be expected has been discussed in papers

by Schiff (1946 a, b) and others (Adams *et al.* 1948, Westendorp *et al.* 1945, Charlton and Breed 1948, Appleyard and Allen-Williams 1949, Mayneord and Martin 1950).

This concentration has in practice both advantages and disadvantages inasmuch as it simplifies the protection problems but provides an automati) limitation of 'field size'. Calculation (Mayneord and Martin 1950) shows that the latter is not serious below, say, 50 MEV., but it may be necessary to insert shaped filters of carbon and metal into the beam to produce a reasonably uniform dose over the area of skin to be irradiated. The most serious disadvantage of these filters is, in our own experience, the loss of output entailed; but with the Kerst betatrons, giving very large outputs, the difficulty does not arise.

There are a number of subsidiary problems in the use of high energy accelerators of the betatron type which must be overcome. One important problem is the loud noise emitted owing to mechanical movement and magnetostriction associated with the powerful alternating current magnets. By careful design of support of the synchrotron and by the use of acoustic tiles on roof and walls of the laboratory we have found that the noise level may be reduced to values (70 db.) at which it is no longer serious to the patient being treated. Indeed, some patients actually prefer the noise as it is an indication, absent from many modern X-ray plants, that they are being treated!

Certainly the most important potential rival in clinical practice to accelerators of the betatron type is the linear accelerator, which in the hands of Fry and his colleagues has become a powerful and reliable source of X-rays and electrons (Fry, Harvie *et al.* 1947, 1948, Newbery 1949). Though the excitation technique is complex, the advantages of accessibility of target and high X-ray output may be decisive, leading to more flexible treatment techniques. Again, under the auspices of the Medical Research Council, short mobile linear accelerators of 4–10 MEV. will become available in a year or two's time, and we look forward to the development of these rival therapeutic sources. Below about 16 MEV. no appreciable activity is induced in tissues, and the problems become essentially an extension of the existing techniques with increased penetration of radiation.

§ 12. PROTECTION

Into the enthusiasm engendered by these fascinating developments and possibilities it is necessary to intrude a grave note of caution. Ionizing radiations have powerful and subtle effects on all living organisms, and those who experiment with them are not exempt. Ionizing radiations are no respectors of persons, however academically qualified. Young scientists, perhaps, particularly require to be reminded continually of the hazards of beams of neutrons, protons, gamma-rays, beta-rays, and we therefore conclude with the briefest consideration of this vast subject.

The deleterious action of ionizing radiations has been known for many years. Indeed, one of the first 'biological experiments' was performed unwittingly by Henri Becquerel, who carried in a pocket one of his earliest radioactive preparations with the result that a small 'burn' of the skin of the abdomen was produced. With the discovery of X-rays and their more general use in medicine the melancholy drama unfolded, until in 1921 protective measures were first drawn up and came into general recognition, though having no legal status. Those

who doubt the reality of the dangers may read them in many places as, for example, in a volume of *Strahlentherapie* (1937) dedicated to x-ray martyrs of all nations.

What, then, are the dangers and how may they be guarded against? In an important report of the International Commission on Radiological Protection drawn up last year (1950) they are listed as:

(1) Superficial injuries.
(2) General effects on the body, particularly the blood and blood-forming organs, e.g. production of anaemia and leukaemias.
(3) The induction of malignant tumours.
(4) Other deleterious effects, including cataract, obesity, impaired fertility, and reduction of life span.
(5) Genetic effects.

Accidents resulting in localized superficial injuries are nowadays usually caused by very large doses, of the order of thousands of röntgens, on a small area, as for example by careless exposure of fingers in the very intense beams from crystallographic equipment where very thin windows very close to the x-ray target are employed. Other accidents have occurred with scattered electron beams (Robbins *et al.* 1946). The direct electron beam from high energy accelerators might, of course, be lethal in a matter of seconds. These effects are obvious, accidents fortunately very rare, and easily guarded against. Probably the worst dangers now arise from beta-rays from containers handled directly.

The general effects are more subtle usually and the result of long-continued exposure at low dosage rates. It must be emphasized that the problem is usually not detection but rather assessment of significant dosage rate level of effect. There is experimental evidence that biological effects, for example changes in the time-scale of development of single organisms, can occur by alteration of the level by irradiation even when that level is of the order of cosmic rays. One of the best known general effects of radiation on man is that on the blood and blood-forming organs, as, for example, the reduction of total number of circulating white cells per cubic millimetre as well as an alteration in the relative proportions of different kinds of cells. Many obscurities persist, and there is, for example, even doubt whether at very low doses the number is not first increased; but this does not invalidate the undoubted usefulness of regular cell counts as a precautionary measure. The effects may also be seen in radiation therapy where necessarily large doses are given, for example in the treatment of malignant disease. The effects are very variable from person to person and there is obvious difficulty in finding out what level of repeated dose (*a*) produces detectable effects, or (*b*) is seriously harmful. The International Commission, reviewing the available evidence obtained on man and animals, came to the conclusion that the maximum permissible amount of radiation should be 0·5 r/week measured on the surface of the body. In the case of high energy beta-rays the maximum permissible exposure is set at the energy equivalent of 1·5 r/week of hard gamma-rays. Different effects of the radiation are often associated with the irradiation of a particular organ or tissue, and in attempting to find some general basis for protective recommendations it is postulated that the figure of 0·3 r/week should not be exceeded for the irradiation of any critical tissue, and it is thought unwise to set for exposure of part of the body, a lower figure than for whole-body irradiation. With respect to neutrons the maximum permissible energy

absorption per gramme exposed to fast neutrons should not be greater than one-tenth of that permitted for high energy quantum radiation.

The induction of malignant tumours in animals by ionizing radiations is a well-known phenomenon. It is exemplified in man in the production of malignancy in over-irradiated hands, but in practice the level of irradiation required is probably higher than for the effects previously considered.

Very little is known with certainty of the genetic effects of radiation in man, but in animals and in a very wide range of species the production of mutations is well known. The significance of these general effects on the population as a whole turns on the fraction of the population which is irradiated since this determines the chance of mating between individuals both of whom are carrying recessive effects. Caution is necessary as it is difficult to forecast the fraction of the population which will be irradiated during the next few years, but it will probably increase in view of the increase of the population engaged in atomic research and industry, as well as the increasing use of x-rays both industrially and medically. In considering such important problems as the disposal of radioactive effluents, and possible contamination of water supplies with small amounts of radioactive materials, it is therefore wise to introduce still further factors of safety until the permitted levels are little above the naturally occurring ionization. Such considerations have been the practice in this country in dealing with these problems.

We cannot here discuss in detail either the maximum permissible amounts of radioactive isotopes in the body or detailed protective measures (see reports by the International Commission on Radiological Protection, 1950). These permissible amounts are normally estimated on the basis of a dose of 0·3 r/week at the critical tissue, calculated as previously indicated in the section on dosimetry. In the case of isotopes like radium, plutonium or strontium, which are deposited in bone, the maximum permissible amounts are based on the observed effects (chiefly tumour production and anaemia) of amounts of radium which have been accidentally (or sometimes deliberately) ingested. The maximum permissible amount of radium fixed in the body is 0·1 μc. Data on the metabolism of radioactive materials are scanty and the International Protection Commission did not feel itself able to make firm recommendations, but has circulated in a Supplement the best authenticated values used in the United States of America, Canada and the United Kingdom as a basis for action and further discussion.

Here we are chiefly concerned to bring to the notice of all those using ionizing radiations, either from high energy accelerators or radioactive isotopes, the need for protective measures. Details of these measures may be found either in the *Report of the International Commission* or, for example, the *Introductory Manual on the Control of Health Hazards from Radioactive Materials* issued by the Medical Research Council (1949). Further information may be found in the writer's report on " Some Applications of Nuclear Physics in Medicine " (Mayneord 1950, pp. 58, 66).

The information upon which our protective measures must be founded is very scanty and often unsatisfactory, and it is greatly to be wished that all research workers in the field would assist by carrying calibrated photographic films and submitting to periodic blood examinations so that data on the correlation of dose and effect could be quickly collected. Indeed, in the writer's view no work with ionizing radiations should be undertaken unless those carrying it out are instructed

in the hazards, are aware at least of the elementary safety measures necessary, and are under proper continuous medical surveillance. Those directing groups of workers have a serious responsibility in this matter.

ACKNOWLEDGMENTS

This report is largely based upon a more detailed account recently published as a Supplement to the *British Journal of Radiology*, The author is much indebted to the British Institute of Radiology for the loan of most of the blocks.

REFERENCES

ADAMS, G. D., ALMY, G. M., DANCOFF, S. M., HANSON, A. O., KERST, D. W., KOCH, H. W., LANZL, E., LANZL, L. H., LAUGHLIN, J. S., QUASTLER, H., RIESEN, D. E., ROBINSON, C. S., and SKAGGS, L. S., 1948, *Amer. J. Roentgenol. Radium Ther.*, **60,** 153.

ALLEN, H., HEMPELMANN, L. H., and NOVACK, N. A., 1945, *Cancer Research*, **5,** 239.

APPLEYARD, R. K., and ALLEN-WILLIAMS, D. J., 1949, *Brit. J. Radiol.*, **22,** 106.

CHAPMAN, E. M., and EVANS, R. D., 1946, *J. Amer. Med. Assoc.*, **131,** 86.

CHARLTON, E. E., and BREED, H. E., 1948, *Amer. J. Roentgenol. Radium Ther.*, **60,** 158.

COHN, W. E., and GREENBERG, D. M., 1939, *J. Biol. Chem.*, **130,** 625.

CORK, J. M., KELLER, H. B., and SAZYNSKI, J., 1949, *Phys. Rev.*, **75,** 1621.

CUTHBERTSON, E. M., and GREENBERG, D. M., 1945, *J. Biol. Chem.*, **160,** 83.

DOBSON, E., GOFMAN, J., JONES, H., KELLY, L., and WALKER, L., 1949, *J. Lab. Clin. Med.*, **34,** 305.

DOUGHERTY, E. C., and LAWRENCE, J. H., 1948, *Advances in Biological and Medical Physics*, Vol. 1 (New York : Academic Press).

ELMER, A. W., 1938, *Iodine Metabolism and Thyroid Function* (New York and London : Oxford University Press).

EVANS, T. C., and QUIMBY, E. H., 1946, *Amer. J. Roentgenol. Radium Ther.*, **55,** 55.

FAILLA, G., 1945, *Amer. J. Roentgenol. Radium Ther.*, **54,** 553.

FLEXNER, L. B., and POHL, H. A., 1940, *Proc. Soc. Exp. Biol. Med.*, **44,** 345; 1941 a, *J. Cell. Comp. Physiol.*, **18,** 49; 1941 b, *Amer. J. Physiol.*, **132,** 594; 1941 c, *J. Biol. Chem.*, **139,** 163; 1941 d, *Amer. J. Physiol.*, **134,** 344.

FLEXNER, L. B., and ROBERTS, R. B., 1939, *Amer. J. Physiol.*, **128,** 154.

FOX, C. L., and KESTON, A. S., 1945, *Surg. Gynec. Obst.*, **80,** 561.

FREUNDLICH, H. F., 1949, *Nature, Lond.*, **164,** 308.

FREUNDLICH, H. F., HAYBITTLE, J. L., and QUICK, R. S., 1950, *Acta Rad.*, **34,** 115.

FRY, D. W., HARVIE, R. B., MULLETT, L. B., and WALKINSHAW, W., 1947, *Nature, Lond.*, **160,** 351; 1948, *Ibid.*, **162,** 859.

GIBSON, J. G., and FINCH, C., 1948, *Advances in Biological and Medical Physics* (quoted by P. F. Hahn), Vol. 1 (New York : Academic Press).

GIBSON, J. G., SELIGMAN, A. M., PEACOCK, W. C., FINE, J., AUB, J. C., and EVANS, R. D., 1947 a, *J. Clin. Invest.*, **26,** 126.

GIBSON, J. G., AUB, J. C., EVANS, R. D., PEACOCK, W. C., IRVINE, J. W., and SACK, T., 1947 b, *J. Clin. Invest.*, **26,** 704.

GIBSON, J. G., EVANS, R. D., AUB, J. C., SACK, T., and PEACOCK, W. C., 1947 c, *J. Clin. Invest.*, **26,** 715.

GRAY, L. H., 1937, *Brit. J. Radiol.*, **10,** 600, 721; 1943, *Proc. Camb. Phil. Soc.*, **40,** 72.

GRAY, L. H., and READ, J., 1939, *Nature, Lond.*, **144,** 439.

GREENBERG, D. M., COPP, D. H., and CUTHBERTSON, E. M., 1943, *J. Biol. Chem.*, **147,** 749.

HAHN, P. F., 1948, *Advances in Biological and Medical Physics*, Vol. 1, p. 287 (New York : Academic Press).

HAHN, P. F., BALE, W. F., LAWRENCE, E. O., and WHIPPLE, G. H., 1938, *J. Amer. Med. Assoc.*, **111,** 2285; 1939, *J. Exper. Med.*, **69,** 739.

HAHN, P. F., BALE, W. F., ROSS, J. F., BALFOUR, W. M., and WHIPPLE, G. H., 1943, *J. Exper. Med.*, **78,** 169.

HAMILTON, J. G., 1942, *Radiology*, **39,** 541.

HAMILTON, J. G., SOLEY, M. H., and EICHORN, K. B., 1940, *Univ. California Publ. in Pharmacol.*, **1**, 339.
HAMILTON, J. G., and SOLEY, M. H., 1940, *Amer. J. Physiol.*, **131**, 135.
HAMILTON, J. G., and STONE, R. S., 1937, *Radiology*, **28**, 178.
HERTZ, S., and ROBERTS, A., 1946, *J. Amer. Med. Assoc.*, **131**, 81.
HERTZ, S., ROBERTS, A., and EVANS, R. D., 1938, *Proc. Soc. Exp. Biol. Med.*, **38**, 510.
HERTZ, S., ROBERTS, A., MEANS, J. H., and EVANS, R. D., 1940, *Amer. J. Physiol.*, **128**, 565.
HEVESY, G., 1948, *Radioactive Indicators* (New York : Interscience Publishers).
INSTITUTE OF PHYSICS, 1950, *Acceleration of Particles to High Energies* (London).
INTERNATIONAL COMMISSION ON RADIOLOGICAL PROTECTION, 1950, Sixth International Congress of Radiology, 45 Lincoln's Inn Fields, London W.C.2.
JOHNS, H. E., DARBY, E. K., WATSON, T. A., and BURKELL, C. C., 1950, *Brit. J. Radiol.*, **23**, 290.
JONES, H. B., CHAIKOFF, I. L., and LAWRENCE, J. H., 1939, *J. Biol. Chem.*, **128**, 631; 1940 a, *Ibid.*, **133**, 319; 1940 b, *Amer. J. Cancer*, **40**, 235; 1940 c, *Ibid.*, **40**, 243.
JONES, H. B., WROBEL, C. J., and LYONS, W. R., 1944, *J. Clin. Invest.*, **23**, 783.
KESTON, A. S., BALL, R. P., FRANTZ, V. K., and PALMER, W. W., 1942, *Science*, **95**, 362.
LAURITSEN, C. C., 1933, *Amer. J. Roentgenol. Radium Ther.*, **30**, 380, 529.
LAWRENCE, J. H., 1949, *J. Amer. Med. Assoc.*, **141**, 13.
LAWRENCE, J. H., and SCOTT, K. G., 1939, *Proc. Soc. Exper. Biol. Med.*, **40**, 694.
LAWRENCE, J. H., SCOTT, K. G., and TUTTLE, L. W., 1939, *New International Clinics*, **3**, 33.
LAWRENCE, J. H., TUTTLE, L. W., SCOTT, K. G., and CONNOR, C. L., 1940, *J. Clin. Invest.*, **19**, 267.
LEBLOND, C. P., 1948, *Advances in Biological and Medical Physics*, Vol. 1 (New York : Academic Press).
LEBLOND, C. P., PUPPEL, I. D., RILEY, E., RADIKE, M., and CURTIS, G. M., 1946, *J. Biol. Chem.*, **162**, 275.
LEITER, L., SEIDLIN, S. M., MARINELLI, L. D., and BAUMANN, E. J., 1946, *J. Clin. Endocrin.*, **6**, 247.
LINDGREN, E., 1944, *Acta Rad.*, **35**, 614.
LOW-BEER, B. V. A., 1946, *Radiology*, **47**, 213; 1947, *Amer. J. Roentgenol. Radium Ther.*, **58**, 4; 1950, *Clinical Use of Radioactive Isotopes* (Springfield, Ill. : Charles C. Thomas).
LOW-BEER, B. V. A., BELL, H. G., MCCORKLE, H. J., STONE, R. S., STEINBACH, H. L., and HILL, W. B., 1946, *Radiology*, **47**, 492.
MARINELLI, L. D., 1942, *Amer. J. Roentgenol. Radium Ther.*, **47**, 210.
MARINELLI, L. D., BRINCKERHOFF, R. F., and HINE, G. J., 1947, *Rev. Mod. Phys.*, **19**, 25.
MARINELLI, L. D., FOOTE, F. W., HILL, R. F., and HOCKER, A. F., 1947, *Amer. J. Roentgenol. Radium Ther.*, **58**, 17.
MAYNEORD, W. V., 1932, *Brit. J. Radiol.*, **5**, 677; 1937, *Acta Int. Union against Cancer*, **2**, 271; 1940, *Brit. J. Radiol.*, **13**, 235; 1942, *Ibid.*, **15**, 100; 1944, *Ibid.*, **17**, 359; 1945, *Ibid.*, **18**, 12; 1946, *Nat. Res. Council of Canada*, Report MM-237; 1950, *Brit. J. Radiol.*, Supplement No. 2.
MAYNEORD, W. V., and BELCHER, E. H., 1950, *Brit. J. Radiol.*, Supplement No. 2, p. 259.
MAYNEORD, W. V., and CIPRIANI, A. J., 1947, *Canad. J. Res.*, **25**, 303.
MAYNEORD, W. V. and CLARKSON, J. R., 1944, *Brit. J. Radiol.*, **17**, 151, 177.
MAYNEORD, W. V., and MARTIN, J. H., 1950, *Brit. J. Radiol.*, Supplement No. 2, p. 212.
MAYNEORD, W. V., MARTIN, J. H., and LAYNE, D. A., 1949, *Nature, Lond.*, **164**, 728.
MAYNEORD, W. V., and ROBERTS, J. E., 1937, *Brit. J. Radiol.*, **10**, 365.
MEDICAL RESEARCH COUNCIL, 1949, *Introductory Manual on the Control of Health Hazards from Radioactive Materials*. (Prepared by the Ministry of Supply, Atomic Energy Research Establishment.)
MEYER, S., and SCHWEIDLER, E., 1927, *Radioaktivität* (Berlin : Teubner).
MITCHELL, J. S., 1947 a, *Brit. J. Radiol.*, **20**, 79; 1947 b, *Ibid.*, **20**, 177; 1947 c, *Ibid.*, **20**, 368.
MORGAREIDGE, K., and MANLY, M. L., 1939, *J. Nutrition*, **18**, 411.
NEWBERY, G. R., 1949, *Brit. J. Radiol.*, **22**, 473.
NICKSON, M., LAMERTON, L. F., and MAYNEORD, W. V., 1949, *Nature, Lond.*, **164**, 613.
OSBORN, S. B., and WRIGHT, H. P., 1949, *Brit. J. Radiol.*, **22**, 110.

POCHIN, E. E., 1950, *Lancet*, July, pp. 41 and 83.
QUIMBY, E. H., and SMITH, B. C., 1944, *Science*, **100,** 175.
RAPER, J. R., 1947, *Radiology*, **49,** 314.
REASER, P. B., and BURCH, G., 1946, *Proc. Soc. Exper. Biol. Med.*, **63,** 543.
REINHARD, E. H., MOORE, C. V., BIERBAUM, O. S., and MOORE, S., 1946, *J. Lab. Clin. Med.*, **31,** 107.
RHOADS, C. P., 1946, *U.S. and U.N. Report Series* 5, *Int. Control of Atomic Energy*, p. 133, Publ. No. 2661.
ROBBINS, L. L., AUB., J. C., COPE, O, COGAN, D. G., LANGOHR, J. L., CLOUD, R. W., and MERRILL, O. E., 1946, *Radiology*, **46,** 1.
ROSS, J. F., and CHAPIN, M. A., 1943, *J. Amer. Med. Assoc.*, **123,** 827.
SALTER, W. T., 1940, *The Endocrine Function of Iodine* (Cambridge, Mass. : Harvard University Press).
SCHIFF, L. I., 1946 a, *Rev. Sci. Instrum.*, **17,** 6; 1946 b, *Phys. Rev.*, **70,** 87.
SEIDLIN, S. M., MARINELLI, L. D., and OSHRY, E., 1946, *J. Amer. Med. Assoc.*, **132,** 838.
SIEVERT, R., 1932, *Acta Rad.*, Supplementum 14.
SMITH, B. C., and QUIMBY, E. H., 1944, *Surg. Gynec. Obst.*, **79,** 142; 1945, *Radiology*, **45,** 335.
SPIERS, F. W., 1946, *Brit. J. Radiol.*, **19,** 52.
STONE, R. S., and LARKIN, J. C., 1942, *Radiology*, **39,** 608.
STONE, R. S., LAWRENCE, J. H., and AEBERSOLD, P. C., 1940, *Radiology*, **35,** 322.
Strahlentherapie, Ehrenbuch der Röntgenologen und Radiologen aller Nationen, Sonderbände zur Strahlentherapie, **22,** 1937.
STREIB, J. F., FOWLER, W. A., and LAURITSEN, C. C., 1941, *Phys. Rev.*, **59,** 253.
TAIT, J. H., 1948, *Atomic Energy Research Establishment Report* No. T/R 231; 1949, *Ibid.*, No. T/R 273; 1950, *Brit. J. Radiol.*, **23,** 282.
TAYLOR, L. S., and SINGER, G., 1940, *Amer. J. Roentgenol. Radium Ther.*, **44,** 428.
THYGESEN, J. E., VIDEBOEK, A., and VILLAUME, I., 1944, *Acta Rad.*, **25,** 305.
TUTTLE, L. W., SCOTT, K. G., and LAWRENCE, J. H., 1939, *Proc. Soc. Exper. Biol. Med.*, **41,** 20.
VEALL, N., 1948, *Brit. J. Radiol.*, **21,** 347.
VISSCHER, M. B., VARCO, R. H., CARR, C. W., DEAN, R. B., and ERICKSON, D., 1944, *Amer. J. Physiol.*, **141,** 488.
WALLACE, D. M., WALTON, R. J., and SINCLAIR, W. K., 1949, *Brit. J. Urol.*, **21,** 357.
WESTENDORP, W. F., and CHARLTON, E. E., 1945, *J. Appl. Phys.*, **16,** 581.
WILSON, R. R., 1946, *Radiology*, **47,** 487.
WRIGHT, H. P., OSBORN, S. B., and EDMONDS, D. G., 1948, *Lancet*, November, p. 767.